



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 166704

TO: Nita M Minnifield  
Location: REM/3C01/3C18  
Art Unit: 1645  
Thursday, September 29, 2005  
Case Serial Number: 09/298523

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen 1a69  
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

### Search Notes

*Reviewed  
9/05*

**THIS PAGE IS BLANK**

STIC-Biotech/ChemLib

166904

From: Minnifield, Nita  
Sent: Friday, September 23, 2005 12:37 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search

09/298523

STIC

Please do a commercial and interference sequence search on SEQ  
ID NO: 13 of this application.

Please provide a paper copy.

Thanks,  
Minnifield  
71976  
Art Unit 1645  
Office REM-3C01  
Mailbox REM-3C18  
571-272-0860

RECEIVED  
SEP 23 2005  
TECH/CHEM. DIVISION  
(STIC)

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2005, 18:42:50 ; Search time 173 Seconds  
(without alignments)  
1206.960 Million cell updates/sec

Title: US-09-298-523D-13  
Perfect score: 2546  
Sequence: 1 MFASKSRKVHYSIRKFSVG.....NRLTQQQPKTEKPAQPSGP 513

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1826554 seqs, 407025358 residues

Total number of hits satisfying chosen parameters: 1826554

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2546	100.0	693	10	US-09-769-787-185
2	2546	100.0	693	10	US-09-769-787-185
3	2546	100.0	693	17	US-10-333-120A-10
4	2516	98.8	511	9	US-09-748-875-13
5	2516	98.8	511	10	US-09-298-523B-13
6	2516	98.8	511	18	US-10-341-201-13
7	2505	98.4	513	9	US-09-748-875-12
8	2505	98.4	513	10	US-09-298-523B-12
9	2505	98.4	513	18	US-10-341-201-12
10	2366	92.9	655	10	US-09-056-019-2
11			655	16	US-10-751-702-2

12	2268	89.1	460	10	US-09-056-019-38	Sequence 38, Appl
13	2268	89.1	460	16	US-10-751-702-38	Sequence 38, Appl
14	2251	88.4	459	10	US-09-056-019-39	Sequence 39, Appl
15	2251	88.4	459	16	US-10-751-702-39	Sequence 39, Appl
16	2210	86.8	446	14	US-10-254-995-9	Sequence 9, Appl
17	2210	86.8	446	20	US-11-062-080-9	Sequence 9, Appl
18	2193	86.1	446	14	US-10-254-995-6	Sequence 6, Appl
19	2193	86.1	446	20	US-11-062-080-6	Sequence 6, Appl
20	2119	83.2	451	9	US-09-748-875-67	Sequence 67, Appl
21	2119	83.2	451	10	US-09-298-523B-67	Sequence 67, Appl
22	2119	83.2	451	18	US-10-341-201-67	Sequence 24, Appl
23	2117	83.2	428	10	US-09-056-019-24	Sequence 24, Appl
24	2117	83.2	428	16	US-10-751-702-24	Sequence 1, Appl
25	1995	78.4	406	10	US-09-056-019-1	Sequence 1, Appl
26	1995	78.4	406	16	US-10-751-702-1	Sequence 9, Appl
27	1796.5	70.6	487	15	US-10-333-120A-9	Sequence 9, Appl
28	1787.5	70.2	487	9	US-09-748-875-9	Sequence 9, Appl
29	1787.5	70.2	487	10	US-09-298-523B-9	Sequence 9, Appl
30	1787.5	70.2	487	18	US-10-341-201-9	Sequence 66, Appl
31	1771.5	69.6	487	9	US-09-748-875-66	Sequence 66, Appl
32	1771.5	69.6	487	10	US-09-298-523B-66	Sequence 66, Appl
33	1771.5	69.6	487	18	US-10-341-201-66	Sequence 6, Appl
34	1756	69.0	701	15	US-10-282-122A-74228	Sequence 74228, A
35	1752	68.8	701	15	US-10-333-120A-7	Sequence 7, Appl
36	1744	68.5	581	9	US-09-748-875-56	Sequence 56, Appl
37	1744	68.5	581	10	US-09-298-523B-56	Sequence 56, Appl
38	1744	68.5	581	18	US-10-341-201-56	Sequence 56, Appl
39	1741	68.4	483	9	US-09-748-875-10	Sequence 10, Appl
40	1741	68.4	483	10	US-09-298-523B-10	Sequence 10, Appl
41	1741	68.4	483	18	US-10-341-201-10	Sequence 10, Appl
42	1726	67.8	481	9	US-09-748-875-6	Sequence 6, Appl
43	1726	67.8	481	10	US-09-298-523B-6	Sequence 6, Appl
44	1726	67.8	481	18	US-10-341-201-6	Sequence 6, Appl
45	1705.5	67.0	496	9	US-09-748-875-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-769-787-185  
; Sequence 185, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388.  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 185  
; LENGTH: 693  
; TYPE: PPT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-185

Query Match	100.0%;	Score 2546;	DB 10;	Length 693;
Best Local Similarity	100.0%;	Pred. No. 1.7e-107;		
Matches 513;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MFASKSRKVHYSIRKFSVGVS	VVVASLVMGVS	VVHATENEGATQVPTSSNRANESQAEQ 60
Db	1	MFASKSRKVHYSIRKFSVGVS	VVVASLVMGVS	VVHATENEGATQVPTSSNRANESQAEQ 60
Qy	61	GEOPKLLSEDKARKEVEYKTKVGSYAKSTKKRHTITVALNNELNNTKVEYLNKIV 120		

Db 61 GEOPKLDSDRKARKEVEEVKKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIV 120  
QY 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
Db 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
QY 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
QY 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300  
Db 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300  
QY 301 ATPDKKENDAKSSDSSVGEETLPSLKPCKKVAEAKKVEEAKKKAEDDQKEEDRRNYP 360  
Db 301 ATPDKKENDAKSSDSSVGEETLPSLKPCKKVAEAKKVEEAKKKAEDDQKEEDRRNYP 360  
QY 361 NTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAIAEVESKKAETRLKIKT 420  
Db 361 NTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAIAEVESKKAETRLKIKT 420  
QY 421 DRKAEAEAKRAAEEDKVKKEPAEQPPAPAPAKAEKPAPAPKDPENPAEQPKAEKPADQ 480  
Db 421 DRKAEAEAKRAAEEDKVKKEPAEQPPAPAPAKAEKPAPAPKDPENPAEQPKAEKPADQ 480  
QY 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513  
Db 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513

## RESULT 2

US-10-333-120A-10

; Sequence 10, Application US/10333120A  
; Publication No. US20040091495A1  
; GENERAL INFORMATION:  
; APPLICANT: Janulczyk, Robert  
; APPLICANT: Iannelli, Francesco  
; APPLICANT: Sjoelhm, Anders  
; APPLICANT: Pozzi, Gianni  
; APPLICANT: Bjorck, Lars  
; TITLE OF INVENTION: PROTEIN  
; CURRENT APPLICATION NUMBER: 13519PCTUS  
; FILE REFERENCE: 13519PCTUS  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: PCT/E901/08409  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: SE 0002728-4  
; PRIOR FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-333-120A-10

Query Match 100.0%; Score 2546; DB 15; Length 693;  
Best Local Similarity 100.0%; Pred. No. 1.7e-107;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
QY 61 GEOPKLDSDRKARKEVEEVKKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIV 120  
Db 61 GEOPKLDSDRKARKEVEEVKKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIV 120  
QY 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
Db 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
QY 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240

QY 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
QY 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300  
Db 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300  
QY 301 ATPDKKENDAKSSDSSVGEETLPSLKPCKKVAEAKKVEEAKKKAEDDQKEEDRRNYP 360  
Db 301 ATPDKKENDAKSSDSSVGEETLPSLKPCKKVAEAKKVEEAKKKAEDDQKEEDRRNYP 360  
QY 361 NTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAIAEVESKKAETRLKIKT 420  
Db 361 NTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAIAEVESKKAETRLKIKT 420  
QY 421 DRKAEAEAKRAAEEDKVKKEPAEQPPAPAPAKAEKPAPAPKDPENPAEQPKAEKPADQ 480  
Db 421 DRKAEAEAKRAAEEDKVKKEPAEQPPAPAPAKAEKPAPAPKDPENPAEQPKAEKPADQ 480  
QY 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513  
Db 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513

## RESULT 3

US-10-472-928-4598

; Sequence 4598, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026926WO  
; CURRENT APPLICATION NUMBER: US/10/472,928  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: GB-0107658.7  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 4598  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; OTHER INFORMATION: choline binding protein A (cbpA)  
; OTHER INFORMATION: Cellular location: outside  
; OTHER INFORMATION: Feature of note: WY motif  
; OTHER INFORMATION: Similar to strain R6 sequence 15904036 (0.E+01)  
US-10-472-928-4598

Query Match 100.0%; Score 2546; DB 17; Length 693;  
Best Local Similarity 100.0%; Pred. No. 1.7e-107;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
QY 61 GEOPKLDSDRKARKEVEEVKKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIV 120  
Db 61 GEOPKLDSDRKARKEVEEVKKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIV 120  
QY 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
Db 121 ESTSESQILMMESRSKVDVAVSKEFKEDSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180  
QY 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 181 AEAKKKVVEAEKKAQDQKEEDRRNYPITYTKTLELEIAESDVEVKKAELELVKVKANEPR 240  
QY 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300  
Db 241 DEQIKIQAIAEVESKQAEATRLKKIKTDREABEAEAKRRADAKQGGPKGKRAKGVPGEL 300

```
Db 241 DEQIKQAEAVESKQAEATRLKKIKTDREAEAEAEKBRADAKGQKPKGRAKGVPGEL 300
Qy 301 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 360
Db 301 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 360
Qy 361 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 420
Db 361 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 420
Qy 421 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 480
Db 421 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 480
Qy 481 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 513
Db 481 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 513

RESULT 4
US-09-748-875-13
; Sequence 13, Application US/09748875
; Publication No. US20010016200A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748, 875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298, 523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-13

Query Match 98.8%; Score 2516; DB 9; Length 511;
Best Local Similarity 99.6%; Pred. No. 2.9e-106;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 MFASKSRKRVHSYIRKFSVGVASVVMASLVNGSVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MFASKSRKRVHSYIRKFSVG-ASVVVASLVNGSVVHATENEGATQVPTSSNRANESQAEQ 59
Qy 61 GEQPKLDSERDKARKEVEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 120
Db 60 GEQPKLDSERDKARKEVEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 119
Qy 121 ESTSESQILLMBSRSKVDVAVSFKFEDSSSSSSDSTKPEASDTAKPNKPTPEGKV 180
Db 120 ESTSESQILLMBSRSKVDVAVSFKFEDSSSSSSDSTKPEASDTAKPNKPTPEGKV 179
Qy 181 AEAKKVEAEAKKADQKEEDRRNYPTITYKTLLEIAESDVVEVKKAELVVKVANEPR 240
Db 180 AEAKKVEEA-KKADQKEEDRRNYPTITYKTLLEIAESDVVEVKKAELVVKVANEPR 238
Qy 241 DEQIKQAEAVESKQAEATRLKKIKTDREAEAEAEKBRADAKGQKPKGRAKGVPGEL 300
Db 239 DEQIKQAEAVESKQAEATRLKKIKTDREAEAEAEKBRADAKGQKPKGRAKGVPGEL 298
Qy 301 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 360
Db 299 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 358
Qy 361 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 420
Db 359 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 418
Qy 421 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 480
Db 419 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 478
Qy 481 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 513
Db 479 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 511

RESULT 6
US-10-341-201-13
; Sequence 13, Application US/10341201
; Publication No. US20050196405A1
; GENERAL INFORMATION:
```

```
Db 419 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 478
Qy 481 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 513
Db 479 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 511

RESULT 5
US-09-298-523B-13
; Sequence 13, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298, 523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-13

Query Match 98.8%; Score 2516; DB 10; Length 511;
Best Local Similarity 99.6%; Pred. No. 2.9e-106;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 MFASKSRKRVHSYIRKFSVGVASVVMASLVNGSVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MFASKSRKRVHSYIRKFSVG-ASVVVASLVNGSVVHATENEGATQVPTSSNRANESQAEQ 59
Qy 61 GEQPKLDSERDKARKEVEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 120
Db 60 GEQPKLDSERDKARKEVEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 119
Qy 121 ESTSESQILLMBSRSKVDVAVSFKFEDSSSSSSDSTKPEASDTAKPNKPTPEGKV 180
Db 120 ESTSESQILLMBSRSKVDVAVSFKFEDSSSSSSDSTKPEASDTAKPNKPTPEGKV 179
Qy 181 AEAKKVEAEAKKADQKEEDRRNYPTITYKTLLEIAESDVVEVKKAELVVKVANEPR 240
Db 180 AEAKKVEEA-KKADQKEEDRRNYPTITYKTLLEIAESDVVEVKKAELVVKVANEPR 238
Qy 241 DEQIKQAEAVESKQAEATRLKKIKTDREAEAEAEKBRADAKGQKPKGRAKGVPGEL 300
Db 239 DEQIKQAEAVESKQAEATRLKKIKTDREAEAEAEKBRADAKGQKPKGRAKGVPGEL 298
Qy 301 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 360
Db 299 ATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAEKKVEPAKKAEDQKEEDRRNYPT 358
Qy 361 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 420
Db 359 NTYKTLLEIAESDVVEVKKAELVKEEAKPRNEEKVKQAKAEVSKKAEATRLLEIKT 418
Qy 421 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 480
Db 419 DRKKAEEBAKKAABEDVKKEPAEQPAPAPAKAEKPAPAPKPNPAEQPKAEKPADOQ 478
Qy 481 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 513
Db 479 AEEDYARRSEBEYNRLTQQQPPKTEKPAQSTP 511

RESULT 6
US-10-341-201-13
; Sequence 13, Application US/10341201
; Publication No. US20050196405A1
; GENERAL INFORMATION:
```

; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/10/341,201  
; CURRENT FILING DATE: 2003-01-13  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-341-201-13

Query Match 98.8%; Score 2516; DB 18; Length 511;  
Best Local Similarity 99.6%; Pred. No. 2.9e-106;  
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
Qy 1 MFASKSERKVVHYSIRKFSVGVASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVVHYSIRKFSVGV-ASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59  
Qy 61 GEQPKLDSERDKARKEVEEYVKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 120  
Db 60 GEQPKLDSERDKARKEVEEYVKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 119  
Qy 121 ESTSESOLQILMMESRSKVDKPEKSDSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 180  
Db 120 ESTSESOLQILMMESRSKVDKPEKSDSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 179  
Qy 181 AEAKKVEEA- KKAQDQKEEDRRNYPTITYKTLEIAESDVEVKAELELVKYKANEPR 240  
Db 180 AEAKKVEEA- KKAQDQKEEDRRNYPTITYKTLEIAESDVEVKAELELVKYKANEPR 238  
Qy 241 DEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRRADAKBQKPKGRKRGVPGEL 300  
Db 239 DEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRRADAKBQKPKGRKRGVPGEL 298  
Qy 301 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPT 360  
Db 299 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPT 358  
Qy 361 NTYKTLEIAESDVEVKAELELVKEEAKPRNEEKVKQAEVSKAEATRLKIKT 420  
Db 359 NTYKTLEIAESDVEVKAELELVKEEAKPRNEEKVKQAEVSKAEATRLKIKT 418  
Qy 421 DRKAEAEAKKAAEDKVKKEPAEQPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 480  
Db 419 DRKAEAEAKKAAEDKVKKEPAEQPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 478  
Qy 481 AEEDYARRSEENRLTQQQPPKTEKPAQSTP 513  
Db 479 AEEDYARRSEENRLTQQQPPKTEKPAQSTP 511

RESULT 7  
US-09-748-875-12  
; Sequence 12, Application US/09748875  
; Publication No. US20010016200A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/748,875  
; CURRENT FILING DATE: 2000-12-26  
; PRIOR FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae  
US-09-748-875-12  
Query Match 98.4%; Score 2505; DB 9; Length 513;  
Best Local Similarity 99.2%; Pred. No. 9.2e-106;  
Matches 509; Conservative 1; Mismatches 1; Indels 2; Gaps 2;  
Qy 1 MFASKSERKVVHYSIRKFSVGVASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVVHYSIRKFSVGV-ASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59  
Qy 61 GEQPKLDSERDKARKEVEEYVKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 120  
Db 60 GEQPKLDSERDKARKEVEEYVKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 119  
Qy 121 ESTSESOLQILMMESRSKVDKPEKSDSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 180  
Db 120 ESTSESOLQILMMESRSKVDKPEKSDSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 179  
Qy 181 AEAKKVEEA- KKAQDQKEEDRRNYPTITYKTLEIAESDVEVKAELELVKYKANEPR 240  
Db 180 AEAKKVEEA- KKAQDQKEEDRRNYPTITYKTLEIAESDVEVKAELELVKYKANEPR 238  
Qy 241 DEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRRADAKBQKPKGRKRGVPGEL 300  
Db 239 DEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRRADAKBQKPKGRKRGVPGEL 298  
Qy 301 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPT 360  
Db 299 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPT 358  
Qy 361 NTYKTLEIAESDVEVKAELELVKEEAKPRNEEKVKQAEVSKAEATRLKIKT 420  
Db 359 NTYKTLEIAESDVEVKAELELVKEEAKPRNEEKVKQAEVSKAEATRLKIKT 418  
Qy 421 DRKAEAEAKKAAEDKVKKEPAEQPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 480  
Db 419 DRKAEAEAKKAAEDKVKKEPAEQPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 478  
Qy 481 AEEDYARRSEENRLTQQQPPKTEKPAQSTP 513  
Db 479 AEEDYARRSEENRLTQQQPPKTEKPAQSTP 511

RESULT 8  
US-09-298-523B-12  
; Sequence 12, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-12

Query Match 98.4%; Score 2505; DB 10; Length 513;  
Best Local Similarity 99.2%; Pred. No. 9.2e-106;  
Matches 509; Conservative 1; Mismatches 1; Indels 2; Gaps 2;  
Qy 1 MFASKSERKVVHYSIRKFSVGVASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVVHYSIRKFSVGV-ASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59  
Qy 61 GEQPKLDSERDKARKEVEEYVKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIV 120



Db 60 GEOPKLLDSERDKARKEVEEYVKIIVGESYAKSTKKRHTITVALVNLNINKNEYLNKIV 119  
Qy 121 ESTSESQILLMESRSKVDKAVSKFEKXSSSSSSSSSTKPEASDTAKPNKPTSPGSKV 180  
Db 120 ESTSESQILLMESRSKVDKAVSKFEKXSSSSSSSSSTKPEASDTAKPNKPTSPGSKV 179  
Qy 181 AEAKKKVEAEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 180 AEAKKKVEAEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKVKANEPR 238  
Qy 241 DEQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKGQKPKGRAGKVGPGEL 300  
Db 239 DEQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKGQKPKGRAGKVGPGEL 298  
Qy 301 ATPDKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKVBEAKKAEADQKEEDRRNYP 360  
Db 299 ATPDKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKVBEAKKAEADQKEEDRRNYP 358  
Qy 361 NTYKTLLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVSKKAEATRLKIKT 420  
Db 359 NTYKTLLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVSKKAEATRLKIKT 418  
Qy 421 DRKKAEEAEAKKAAEEDKVKKEPAEQOPAPAPKAEPAPAPKAPENPAEQPKAEKPADOQ 480  
Db 419 DRKKAEEAEAKKAAEEDKVKKEPAEQOPAPAPKAEPAPAPKAPENPAEQPKAEKPADOQ 478  
Qy 481 AEEDYARRSEBYNRLTQQPPKTEKPAQPSTP 513  
Db 479 AEEDYARRSEBYNRLTQQPPKTEKPAQPSTP 511

## RESULT 9

US-10-341-201-12  
; Sequence 12, Application US/10341201  
; Publication No. US20050196405A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/10/341,201  
; CURRENT FILING DATE: 2003-01-13  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-341-201-12

Query Match 98.4%; Score 2505; DB 18; Length 513;  
Best Local Similarity 99.2%; Pred. No. 9.2e-106;  
Matches 509; Conservative 1; Mismatches 1; Indels 2; Gaps 2;  
Qy 1 MFASKSERKVHYSIRKFSVGVASVVVAVSLVMSVHVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKVHYSIRKFSVGV-ASVVAVSLVMSVHVHATENEGATQVPTSSNRANESQAEQ 59  
Qy 61 GEOPKLLDSERDKARKEVEEYVKIIVGESYAKSTKKRHTITVALVNLNINKNEYLNKIV 120  
Db 60 GEOPKLLDSERDKARKEVEEYVKIIVGESYAKSTKKRHTITVALVNLNINKNEYLNKIV 119  
Qy 121 ESTSESQILLMESRSKVDKAVSKFEKXSSSSSSSSSTKPEASDTAKPNKPTSPGSKV 180  
Db 120 ESTSESQILLMESRSKVDKAVSKFEKXSSSSSSSSSTKPEASDTAKPNKPTSPGSKV 179  
Qy 181 AEAKKKVEAEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKVKANEPR 240  
Db 180 AEAKKKVEAEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKVKANEPR 238  
Qy 241 DEQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKGQKPKGRAGKVGPGEL 300  
Db 239 DEQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKGQKPKGRAGKVGPGEL 298

Qy 301 ATPDKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKVBEAKKAEADQKEEDRRNYP 360  
Db 299 ATPDKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKVBEAKKAEADQKEEDRRNYP 358  
Qy 361 NTYKTLLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVSKKAEATRLKIKT 420  
Db 359 NTYKTLLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVSKKAEATRLKIKT 418  
Qy 421 DRKKAEEAEAKKAAEEDKVKKEPAEQOPAPAPKAEPAPAPKAPENPAEQPKAEKPADOQ 480  
Db 419 DRKKAEEAEAKKAAEEDKVKKEPAEQOPAPAPKAEPAPAPKAPENPAEQPKAEKPADOQ 478  
Qy 481 AEEDYARRSEBYNRLTQQPPKTEKPAQPSTP 513  
Db 479 AEEDYARRSEBYNRLTQQPPKTEKPAQPSTP 511

## RESULT 10

US-09-056-019-2  
; Sequence 2, Application US/09056019A  
; Publication No. US20030096950A1  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wizemann, Theresa  
; APPLICANT: Masure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Koenig, Scott  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; TITLE OF INVENTION: THEREFROM AND USES THEREOF  
; FILE REFERENCE: 1340-1-017  
; CURRENT APPLICATION NUMBER: US/09/056.019A  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019-2

Query Match 92.9%; Score 2366; DB 10; Length 655;  
Best Local Similarity 100.0%; Pred. No. 2.2e-99;  
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 ENEGATQVPTSSNRANESQAEQEQPKLLDSERDKARKEVEEYVKIIVGESYAKSTKKRH 98  
Db 1 ENEGATQVPTSSNRANESQAEQEQPKLLDSERDKARKEVEEYVKIIVGESYAKSTKKRH 60  
Qy 99 TITVALVNLNINKNEYLNKIVESTSESQILLMESRSKVDEAVSKFEDSSSSSSSDS 158  
Db 61 TITVALVNLNINKNEYLNKIVESTSESQILLMESRSKVDEAVSKFEDSSSSSSSDS 120  
Qy 159 STKPEASDTAKPNKPTSPGSKVBAEKVBEAKKADQKEEDRRNYPITYKTLELEIA 218  
Db 121 STKPEASDTAKPNKPTSPGSKVBAEKVBEAKKADQKEEDRRNYPITYKTLELEIA 180  
Qy 219 ESDVEVKKAELELVKVKANEPRDQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAEAKR 278  
Db 181 ESDVEVKKAELELVKVKANEPRDQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAEAKR 240  
Qy 279 RADAKGQKPKGRAGKVGPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKPKVAEAK 338  
Db 241 RADAKGQKPKGRAGKVGPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKPKVAEAK 300  
Qy 339 KVBEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKV 398  
Db 301 KVBEAKKADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKV 360  
Qy 399 KQAKAEVSKKAEATRLKIKTDREAEAEAKRKAEEEDKVKKEPAEQOPAPAPKAEP 458  
Db 361 KQAKAEVSKKAEATRLKIKTDREAEAEAKRKAEEEDKVKKEPAEQOPAPAPKAEP 420

```
Qy 459 APAPKPNPAQPKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSPT 513
|||||
Db 421 APAPKPNPAQPKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSPT 475

RESULT 11
US-10-751-702-2
; Sequence 2, Application US/10751702
; Publication No. US20040120966A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 5853-2
; CURRENT APPLICATION NUMBER: US/10751,702
; PRIOR FILING DATE: 2004-01-05
; PRIOR APPLICATION NUMBER: US/09/056,019B
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-751-702-2

Query Match 92.9%; Score 2366; DB 16; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.2e-99;
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 ENEGATQVPTSSNRANESQAEQGGQPKLDSERDKARKEVEEYVKTI VGESYAKSTKKRH 98
|||||
Db 1 ENEGATQVPTSSNRANESQAEQGGQPKLDSERDKARKEVEEYVKTI VGESYAKSTKKRH 60

Qy 99 TITVALVNLNIIKNEYLKIVESTSESQILMMESRSKVDEAVSFKDSSSSSSDS 158
|||||
Db 61 TITVALVNLNIIKNEYLKIVESTSESQILMMESRSKVDEAVSFKDSSSSSSDS 120

Qy 159 STKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKKAQDQKEEDRRNPTTITTKTLELEIA 218
|||||
Db 121 STKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKKAQDQKEEDRRNPTTITTKTLELEIA 180

Qy 219 ESDVEVVKAELELVKVKANEPDRDQKI KQAEAEVESKQAEATRLKKIKTDREAEAEAKR 278
|||||
Db 181 ESDVEVVKAELELVKVKANEPDRDQKI KQAEAEVESKQAEATRLKKIKTDREAEAEAKR 240

Qy 279 RADAKEGKPKGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKKVAEAEK 338
|||||
Db 241 RADAKEGKPKGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKKVAEAEK 300

Qy 339 KVEEAKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVVKAELELVKERAKPRNEEKV 398
|||||
Db 301 KVEEAKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVVKAELELVKERAKPRNEEKV 360

Qy 399 KQAEAEVESKKAETRLLEKI KTDREAEAEAKRKAEEEDVKKEKPAEQPPAPAKAEKP 458
|||||
Db 361 KQAEAEVESKKAETRLLEKI KTDREAEAEAKRKAEEEDVKKEKPAEQPPAPAKAEKP 420

Qy 459 APAPKPNPAQPKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSPT 513
|||||
Db 421 APAPKPNPAQPKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSPT 475

RESULT 12
US-09-056-019-38
; Sequence 38, Application US/09056019A
; Publication No. US20030096950A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match 89.1%; Score 2268; DB 10; Length 460;
Best Local Similarity 99.8%; Pred. No. 4.2e-95;
Matches 458; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 25 VVASLVGMSVVHATENEGATQVPTSSNRANESQAEQGGQPKLDSERDKARKEVEEYVK 84
|||||
Db 2 IVASLVGMSVVHATENEGATQVPTSSNRANESQAEQGGQPKLDSERDKARKEVEEYVK 61

Qy 85 IVGESYAKSTKKGHTTITVALVNLNIIKNEYLKIVESTSESQILMMESRSKVDEAVS 144
|||||
Db 62 IVGESYAKSTKKGHTTITVALVNLNIIKNEYLKIVESTSESQILMMESRSKVDEAVS 121

Qy 145 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKKAQDQKEEDRRN 204
|||||
Db 122 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKKAQDQKEEDRRN 181

Qy 205 YPTITTKTLELEIAESDVEVVKAELELVKVKANEPDRDQKI KQAEAEVESKQAEATRLKK 264
|||||
Db 182 YPTITTKTLELEIAESDVEVVKAELELVKVKANEPDRDQKI KQAEAEVESKQAEATRLKK 241

Qy 265 IKTDREAEAEAKRADAKEQGGPKGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPS 324
|||||
Db 242 IKTDREAEAEAKRADAKEQGGPKGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPS 301

Qy 325 PSLKPEKKVAEAEKKEEAKKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVVKAELEL 384
|||||
Db 302 PSLKPEKKVAEAEKKEEAKKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVVKAELEL 361

Qy 385 VKEEAKEPRNEEKVKQAEAEVESKKAETRLLEKI KTDREAEAEAKKAAEEDVKKEKPA 444
|||||
Db 362 VKEEAKEPRNEEKVKQAEAEVESKKAETRLLEKI KTDREAEAEAKKAAEEDVKKEKPA 421

Qy 445 EQQPAPAPKAEKPAKPAKPNPAEQPKAEKPADQQAEE 483
|||||
Db 422 EQQPAPAPKAEKPAKPAKPNPAEQPKAEKPADQQAEE 460

RESULT 13
US-10-751-702-38
; Sequence 38, Application US/10751702
; Publication No. US20040120966A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 5853-2
; CURRENT APPLICATION NUMBER: US/10751,702
; CURRENT FILING DATE: 2004-01-05
```

```
; PRIOR APPLICATION NUMBER: US/09/056,019B
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-751-702-38

Query Match      89.1%; Score 2268; DB 16; Length 460;
Best Local Similarity 99.8%; Pred. No. 4.2e-95;
Matches 458; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 25 VVASLVMSGVVHATENEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 84
   :|||||
DB 2 IVASLVMSGVVHATENEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 61
   :|||||
QY 85 IVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIVESTSESQLOILMMESRSKVDEAVS 144
   :|||||
DB 61 IVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIVESTSESQLOILMMESRSKVDEAVS 120
   :|||||
QY 145 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKKAQKQKEEDRRN 204
   :|||||
DB 121 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKKAQKQKEEDRRN 180
   :|||||
QY 205 YPTITYKTLELEIAESDVEVKKAELVVKVANEPRDEQIKQAQAEVESKQAEATRLKK 264
   :|||||
DB 181 YPTITYKTLELEIAESDVEVKKAELVVKVANEPRDEQIKQAQAEVESKQAEATRLKK 240
   :|||||
QY 265 IKTDREAEAEERAKRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSVSGEETLPS 324
   :|||||
DB 241 IKTDREAEAEERAKRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSVSGEETLPS 300
   :|||||
QY 325 PSLKPEKKVAEAEKKVEAEKKAQKQKEEDRRNYPNTYKTLELEIAESDVEVKKAEL 384
   :|||||
DB 301 PSLKPEKKVAEAEKKVEAEKKAQKQKEEDRRNYPNTYKTLELEIAESDVEVKKAEL 360
   :|||||
QY 385 VKEEAKPRNEBKVKQAQAEVESKKAATRLBKIKTDRKKAEEAEAKRKAABEDKVKKEPA 444
   :|||||
DB 361 VKEEAKPRNEBKVKQAQAEVESKKAATRLBKIKTDRKKAEEAEAKRKAABEDKVKKEPA 420
   :|||||
QY 445 EQQPAPAPKAPKAPAPKAPENPAPQPKAEKPADQQAEE 483
   :|||||
DB 421 EQQPAPAPKAPKAPAPKAPENPAPQPKAEKPADQQAEE 459
   :|||||

RESULT 15
US-10-751-702-39
; Sequence 39, Application US/10751702
; Publication No. US20040120966A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wisemann, Theresa M.
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 5853-2
; CURRENT APPLICATION NUMBER: US/10/751,702
; CURRENT FILING DATE: 2004-01-05
; PRIOR APPLICATION NUMBER: US/09/056,019B
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-751-702-39

Query Match      88.4%; Score 2251; DB 16; Length 459;
Best Local Similarity 98.9%; Pred. No. 2.5e-94;
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 25 VVASLVMSGVVHATENEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 84
   :|||||
DB 1 IVASLVMSGVVHATENEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 60
   :|||||
QY 85 IVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIVESTSESQLOILMMESRSKVDEAVS 144
   :|||||
DB 61 IVGESYAKSTKKRHTITVALVNLNINIKNEYLNKIVESTSESQLOILMMESRSKVDEAVS 120
   :|||||

Query Match      88.4%; Score 2251; DB 10; Length 459;
Best Local Similarity 98.9%; Pred. No. 2.5e-94;

; PRIOR APPLICATION NUMBER: US/09/056,019A
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-39

Query Match      88.4%; Score 2251; DB 10; Length 459;
Best Local Similarity 98.9%; Pred. No. 2.5e-94;
```

QY	145	KPEKSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKKVEEAEKAKDKQKEEDRRN	204
Db	121		
QY	205	YPTIYKTLLELEIAESDVEVKAELELVKVKANEPRDQKIQAEEVESKQAEATRLKK	264
Db	181		
QY	265	IKTDREAEAEAKRRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEEITLPS	324
Db	241		
QY	325	PSLKPEKKVAEAKKVEEAKKAEQKQKEEDRRNPTNTYKTLLELEIAESDVEVKAELEL	384
Db	301		
QY	385	VKEEAKPRNEEKVKQAKAEVESKKAETRLKIKTDRKKAEEBAKKAABEDVKKEKPA	444
Db	361		
QY	445	EQQPAPAPKAEKPAPAPKPNPAPAEQPKAEKPADOQAAE	483
Db	421	EQQPAPAPKTEKPAPAPKPNPAPAEQPKAEKPADOQAAE	459

Search completed: September 28, 2005, 18:56:50  
Job time : 176 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2005, 18:33:49 ; Search time 43 Seconds  
(without alignments)  
890.581 Million cell updates/sec

Title: US-09-298-523D-13

Perfect score: 2546

Sequence: 1 MFASKSRKRVHSIRKFSVG.....NRLTQQQPPKTEKPAQSTP 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2210	86.8	446	4	US-09-286-981B-9
2	2193	86.1	446	4	US-09-286-981B-6
3	1754	68.9	564	3	US-09-308-022-6
4	1524	59.9	406	4	US-09-286-981B-18
5	1517.5	59.6	631	3	US-08-847-065-25
6	1517.5	59.6	631	4	US-09-829-382-25
7	1467	57.6	426	4	US-09-286-981B-12
8	1459	57.3	425	4	US-09-286-981B-13
9	1458	57.3	424	4	US-09-286-981B-14
10	1452	57.0	428	4	US-09-286-981B-7
11	1443.5	56.7	425	4	US-09-286-981B-11
12	1428	56.1	412	4	US-09-286-981B-17
13	1417	55.7	414	4	US-09-286-981B-16
14	1416	55.6	419	4	US-09-286-981B-15
15	1413	55.5	414	4	US-09-286-981B-10
16	1410.5	55.4	431	4	US-09-286-981B-3
17	1373	53.9	453	3	US-08-961-083-38
18	1373	53.9	453	4	US-09-536-784-38
19	1306.5	51.3	413	4	US-09-286-981B-5
20	1233	48.4	864	4	US-08-714-741-40
21	1168.5	45.9	588	4	US-08-714-741-42
22	915.5	36.0	605	4	US-08-714-741-46
23	904.5	35.5	222	4	US-09-107-433-4344
24	879.5	34.5	216	4	US-09-583-110-4344
25	854	33.5	8991	4	US-08-714-741-32
26	775.5	30.5	1231	4	US-08-714-741-41
27	770	30.2	275	4	US-09-583-110-4347

28	757.5	29.8	623	4	US-08-714-741-47	Sequence 47, Appl
29	714	28.0	251	4	US-09-286-981B-4	Sequence 4, Appl
30	657	25.8	233	4	US-09-107-433-3526	Sequence 3526, Ap
31	636	25.0	142	3	US-08-847-065-21	Sequence 21, Appl
32	636	25.0	142	4	US-08-829-382-21	Sequence 21, Appl
33	595	23.4	131	4	US-08-529-055-56	Sequence 56, Appl
34	576.5	22.6	128	4	US-08-529-055-57	Sequence 57, Appl
35	561.5	22.1	482	4	US-09-107-433-4705	Sequence 4705, Ap
36	560	22.0	219	4	US-09-286-981B-8	Sequence 8, Appl
37	542	21.3	114	4	US-09-286-981B-19	Sequence 19, Appl
38	541	21.2	110	3	US-08-961-083-102	Sequence 102, App
39	541	21.2	110	4	US-09-536-784-102	Sequence 102, App
40	497	19.5	695	1	US-08-127-499A-23	Sequence 23, Appl
41	497	19.5	695	1	US-08-482-847-23	Sequence 23, Appl
42	496	19.5	619	1	US-08-465-746-2	Sequence 2, Appl
43	496	19.5	619	1	US-08-214-164-2	Sequence 2, Appl
44	496	19.5	619	2	US-08-467-852A-3	Sequence 3, Appl
45	496	19.5	619	2	US-08-246-636-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-286-981B-9  
; Sequence 9, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286.981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-9

Query Match		86.8%	Score 2210;	DB 4;	Length 446;
Best Local Similarity		100.0%;	Pred. No. 2.9e-145;		
Matches 446;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	38	TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEYVKIVGESYAKTKKR	97		
Db	1	TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEYVKIVGESYAKTKKR	60		
QY	98	HTITVALNNELNINKYLNKIVESTSSQQLMMESRSKVDEAVSKFEKDSSSSSSD	157		
Db	61	HTITVALNNELNINKYLNKIVESTSSQQLMMESRSKVDEAVSKFEKDSSSSSSD	120		
QY	158	SSTKPEASDTAKPNKPTPEGKVAEAKKKVKEAEKAKDKQEKEDRRNYPTTYKTLELEI	217		
Db	121	SSTKPEASDTAKPNKPTPEGKVAEAKKKVKEAEKAKDKQEKEDRRNYPTTYKTLELEI	180		
QY	218	AESDVEVKAELELVKVKANPRDQKIKQAEAVESQAEATRLKKIKTDRREEAEAEAK	277		
Db	181	AESDVEVKAELELVKVKANPRDQKIKQAEAVESQAEATRLKKIKTDRREEAEAEAK	240		
QY	278	PRADAKEGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSLSLKEKKVAEAE	337		
Db	241	PRADAKEGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSLSLKEKKVAEAE	300		
QY	338	KKVEAKKKAEDQEKEDRRNYPTTYKTLELEIAESDVEVKAELELVKKEAEKPRNEEK	397		
Db	301	KKVEAKKKAEDQEKEDRRNYPTTYKTLELEIAESDVEVKAELELVKKEAEKPRNEEK	360		



Db 259 KRAKRGVPGELATPDKENDAKSSDSSVGBETLPSLSKSGKVAEAKKVEAEKAK 318  
Qy 349 DQKEDRRNYPNTYKTLLEIAESDVEVKKAELELVKEEAKPRNEKVKQAQAEVSK 408  
Db 319 DQKEDRRNYPNTYKTLLEIAESDVEVKKAELELVKEEAKPRNEKVKQAQAEVSK 378  
Qy 409 KAEATRELEKIKTDRKAEAEAKKAAEEDKVEKPAEQOPAPAPKAPKAPAPENPA 468  
Db 379 KAEATRELEKIKTDRKAEAEAKKAAEEDKVEKPAEQOPAPAPKAPKAPAPENPA 436  
Qy 469 EQPKAEKPADQQAEDYARRSEEEYNRLTQOPPKTEKPAQPS 513  
Db 437 EQPKAEKPADQQAEDYARRSEEEYNRLTQOPPKTEKPAQPS 481

## RESULT 4

US-09-286-981B-18  
; Sequence 18, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wisemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-18

Query Match 59.9%; Score 1524; DB 4; Length 406;  
Best Local Similarity 73.0%; Pred. No. 7.6e-98;  
Matches 327; Conservative 29; Mismatches 46; Indels 46; Gaps 7;  
Qy 38 TENEGATQVPTSSNRANESQAEQEQPKLSDERDKARKEVEYVKLVGESYAKTKR 97  
Db 1 TENEGTQAPTSSNRNGESQA-----EHMKAAQVDEYIEKML-----QLORRK 44  
Qy 98 HTITVALVNLNINIKNEYLNKIVESTESQQLIMMESRSKVDEAVSKFEKDSSSSSSD 157  
Db 45 HTQNVGLLTGLGAIKTEYLRGLSVSKSTAE-LPSEIKELTAAPKQPK-----94  
Qy 158 SSTKPEASDTAKPNKPTPEGKVAEAKKVAEAKKAKDQKEEDRRNYPNTYKTLLEI 217  
Db 95 -----DTLKPEK-----KVAEAEKKVAEAKKAKDQKEEDRRNYPNTYKTLLEI 140  
Qy 218 AESDVEVKKAELELVKVKANPREOKTKQAEAVESKQAEATRLKIKTDRREAESEAK 277  
Db 141 AESDVEVKKAELELVKVKANPREOKTKQAEAVESKQAEATRLKIKTDRREAESEAK 200  
Qy 278 RRADAKEQ-----KPKGRKRGVPGELATPDKENDAKSSDSSVGEETLPSPLKPKKV 333  
Db 201 RRADAKEQDESSKRRKSRVKGDLGEQATPDKENDAKSSDSSVGEETLPSPLKPKKV 260  
Qy 334 AEAKKVEEAKKADQKEEDRRNYPNTYKTLLEIAESDVEVKKAELELVKEEAKPR 393  
Db 261 AEAKKVEEADKKAQKEEDRRNYPNTYKTLLEIAESDVEVKKAELELVKEEAKPR 320  
Qy 394 NEEKVKQAEVSKKAEATLEKIKTDRKAEAEKAAEEDKVEKPAEQOPAPAP 453  
Db 321 NEEKVKQAEVSKKAEATLEKIKTDRKAEAEKAAEEDKVEKPAEQOPAPAP 380  
Qy 454 KAEKPAKAPKAPENPAEQPKAEKPADQQA 481  
Db 381 QPEK--PAPKAPENPAEQPKAEKPADQQA 406

## RESULT 5

US-08-847-065-25  
; Sequence 25, Application US/08847065  
; Patent No. 6245335  
; GENERAL INFORMATION:  
; APPLICANT: Masure, H. Robert  
; APPLICANT: Rosenow, Carsten I.  
; APPLICANT: Tuomanen, Elaine  
; APPLICANT: Wisemann, Theresa M.  
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
; TITLE OF INVENTION: ANTI-PNEUMOCOCCAL VACCINES  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/847,065  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-158 ..  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 631 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: N-terminal  
US-08-847-065-25

Query Match 59.6%; Score 1517.5; DB 3; Length 631;  
Best Local Similarity 73.5%; Pred. No. 3.6e-97;  
Matches 316; Conservative 40; Mismatches 35; Indels 39; Gaps 5;  
Qy 95 KCRHTITVALVNLNINIKNEYLNK--IVESTESQQLIMMESRSKVDEAVSKFEKDS 152  
Db 9 RRKHTQVNLNLIKLSAKTKYLRNLNVLSEKSKDEL--SEIKAKLDAAEKFKDT--63  
Qy 153 SSSSDSSTKPEASDTAKPNKPTPEGKVAEAKKVAEAKKAKDQKEEDRRNYPNTYK 212  
Db 64 -----LKPGEKVAEAKKVAEAKKAKDQKEEDRRNYPNTYK 102  
Qy 213 LELEIAESDVEVKKAELELVKVKANPREOKTKQAEAVESKQAEATRLKIKTDRREA 272  
Db 103 LELEIAESDVEVKKAELELVKVKANPREOKTKQAEAVESKQAEATRLKIKTDRREA 162  
Qy 273 EEEAKRADAK-----EQQPKGRKRGVPGELATPDKENDAKSSDSSVGEETLP 323  
Db 163 EEEAKRADAKLKEANVATSDQPKGRKRGVPGELATPDKENDAKSSDSSVGEETLP 222  
Qy 324 SPSLKPKKVAEAEKVEEAKKAKDQKEEDRRNYPNTYKTLLEIAESDVEVKKAELE 383  
Db 223 SSSLKSKKVAEAEKVEEAKKAKDQKEEDRRNYPNTYKTLLEIAESDVEVKKAELE 282  
Qy 384 LVKEEAKPRNEEKVKQAKAEVSKKAEATRLKIKTDRKAEAEKAAEEDKVEKPK 443

Db 283 LVKEAKEPRDEEKIKQAKAKVESKKAETRLNENIKTDKRAEAEAKRAKAAEEDKVKEKP 342  
Qy 444 AEQOPAPAPAKPAPKAPENPAEQPKAKPADQQAEEYARRSEEEYNRLTQQQPPK 503  
Db 343 AEQOPAPATQPEK--PAPKPEKPAEQPKAKETDDQQAEEYARRSEEEYNRLTQQQPPK 400  
Qy 504 TEKPAQSTP 513  
Db 401 TEKPAQSTP 410

RESULT 6  
US-09-829-382-25  
; Sequence 25, Application US/09829382  
; Patent No. 6784164  
; GENERAL INFORMATION:  
; APPLICANT: Masure, H. Robert  
; Rosenow, Carsten I.  
; Tuomanen, Elaine  
; Wizemann, Theresa M.  
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
; ANTI-PNEUMOCOCCAL VACCINES  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/829,382  
; FILING DATE: 09-Apr-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/947,065  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-158 ..  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 631 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: N-terminal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
US-09-829-382-25

Query Match 59.6%; Score 1517.5; DB 4; Length 631;  
Best Local Similarity 73.5%; Pred. No. 3.6e-97;  
Matches 316; Conservative 40; Mismatches 35; Indels 39; Gaps 5;  
Qy 95 KGRHTITVALVNLNNIKNEYLNK--IVESTSESQQLILMMESRSKYDEAVSKPEKDDSS 152  
Db 9 RRKHTQNVALNKLKSAIKTKYLRNLNVLEKSKDELFP--SEIKAKLDAAFKFKDGT-- 63  
Qy 153 SSSSDSSTKPEASDTAKPNKPTPEGEKVAEAKKKVVEAEKAKDQKEDRRNYPTITYKT 212

Db 64 -----LKPGEKVAEAKKKVVEAEKAKDQKEDRRNYPTITYKT 102  
Qy 213 LELEIAESDVEVKAELELVKVKANEPRDROKIKQAEAEVESKQAEATRLKKIKITDREKA 272  
Db 103 LELEIAEFDVKVKEAELELVKEEAKESRNEGTTIKQAEKVESKKAETRLNENIKTDKRA 162  
Qy 273 EBEAKRRADAK-----EQGPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLP 323  
Db 163 EEEAKRKADAKLEANVATSDQCKPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLP 222  
Qy 324 SPSELKPEKKVABAEAKKKVVEAEKAKDQKEDRRNYPTITYKTLELEIAESDVEVKAELE 383  
Db 223 SSSLKSGKKVAEAEKKVVEAEKAKDQKEDRRNYPTITYKTLDLEIAESDVKVKEAELE 282  
Qy 384 LVKEEAKEPNBEKVKQAKAEVESKKAETRLKIKTDKKAEEAKRAAEEDKVKEKP 443  
Db 283 LVKEEAKEPNDEEKIKQAKAKVESKKAETRLNENIKTDKKAEEAKRAAEEDKVKEKP 342  
Qy 444 AEQOPAPAPAKPAPKAPENPAEQPKAKPADQQAEEYARRSEEEYNRLTQQQPPK 503  
Db 343 AEQOPAPATQPEK--PAPKPEKPAEQPKAKETDDQQAEEYARRSEEEYNRLTQQQPPK 400  
Qy 504 TEKPAQSTP 513  
Db 401 TEKPAQSTP 410

RESULT 7  
US-09-286-981B-12  
; Sequence 12, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 426  
; TYPE: PRP  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-12  
Query Match 57.6%; Score 1467; DB 4; Length 426;  
Best Local Similarity 68.4%; Pred. No. 7.1e-94;  
Matches 318; Conservative 42; Mismatches 47; Indels 58; Gaps 8;  
Qy 38 TENEGATQVPTSSNRANESQAEQGEQPKLDSRDKARKEVEYVKIVGESYAKSTKKR 97  
Db 1 TEKEVTTQVATSSNKANKSQ-----TEHMKAAKQVDEYIKKKI-----QLDRRK 44  
Qy 98 HTITVALVNLNNIKNEYLNKIVESTSESQQLILMMESRSKYDEAVSKFEKSSSSSSD 157  
Db 45 HTQNVGLLTGKLVIKTEYHLGLSVSKKSEAE-LPSEIKAKLDAAFEQFKDIT----- 96  
Qy 158 SSTKPEASDTAKPNKPTPEGEKVAEAKKKVVEAEKAKDQKEDRRNYPTITYKTLELEI 217  
Db 97 -----LPTPECKKVAEAEKKVVEAEKAKDQKEDRRNYPTITYKTLELDI 142  
Qy 218 AESDVEVKAELELVKVKANEPRDEOKIKQAEAEVESKQAEATRLKKIKITDREAEAEAK 277  
Db 143 AESDVEVKAELELVKEEAKESRDEKKINQAKAKVENKKAETRLNENIKTDREKA-BAK 201  
Qy 278 RRADAK-----SQGPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK 328  
Db 202 RRADAKLEANVATSDQCKSKRRAKRVLGELATPDKKENDAKSSDSSVGEETLTSPLK 261



QY 329 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 388  
DB 262 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 321  
QY 389 AKEPRNEEKVQKAEVSKAEATRLKIKTDKKA-EEAAKRAAEEDVKKEKPAQP 447  
DB 322 AKESRNEEKIKQVKAKVESKAEATRLKIKTDKKAEEAAKRAAEEDVKKEKPAQP 381  
QY 448 QPAPAPKAEK-----PAPAPKPNPAPQPAEKPADQAAE 483  
DB 382 QPAPAPQEPKTEEPENPAPAPKPNPAPKPAEKPADQAAE 426

## RESULT 8

US-09-286-981B-13  
; Sequence 13, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 425  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-13

Query Match 57.3%; Score 1459; DB 4; Length 425;  
Best Local Similarity 68.3%; Pred. No. 2.5e-93;  
Matches 317; Conservative 42; Mismatches 47; Indels 58; Gaps 8;  
QY 38 TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKKR 97  
DB 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAAKQVDEYIKKKL-----QLDRRK 44  
QY 98 HTITVALVNLNINIKVEYLKIVESTSSQILMMESRSKVDKAEVSKFEDSSSSSSD 157  
DB 45 HTQNVGLLTGLGVIKTEYLHGLSVSKKSEAE-LPSEIKAKLDAAPFQFKDT----- 96  
QY 158 SSTKPEASDTAKPNKPTPEGEKVAEAKKVEAEAKKADOKEDRRNYPNTYKTLLEI 217  
DB 97 -----LPTPEGKKVAEAEKKVEAEAKKADOKEDRRNYPNTYKTLLEI 142  
QY 218 AESDVEVKAELELVKVKANEPDRQKIQAEEVSKAEATRLKIKTDREAEAEAEAK 277  
DB 143 AESDVEVKAELELVKVEAKESRDEKKNQAKAVENKKAATRLKIKTDREKA-EEAK 201  
QY 278 RRADAK-----EQGPKGRAKGVPGELATPDKKENDAKSSDSSVGBETLPSPLK 328  
DB 202 RRADAKQEAENVATSEQDKSKRAKRVFGLATPDKKENDAKSSDSSVGBETLPSPLK 261  
QY 329 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 388  
DB 262 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 321  
QY 389 AKEPRNEEKVQKAEVSKAEATRLKIKTDKKA-EEAAKRAAEEDVKKEKPAQP 447  
DB 322 AKESRNEEKIKQVKAKVESKAEATRLKIKTDKKAEEAAKRAAEEDVKKEKPAQP 381  
QY 448 QPAPAPKAEK-----PAPAPKPNPAPQPAEKPADQAAE 482  
DB 382 QPAPAPQEPKTEEPENPAPAPKPNPAPKPAEKPADQAAE 425

## RESULT 9

US-09-286-981B-14  
; Sequence 14, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 424  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-14

Query Match 57.3%; Score 1458; DB 4; Length 424;  
Best Local Similarity 68.5%; Pred. No. 2.9e-93;  
Matches 317; Conservative 41; Mismatches 47; Indels 58; Gaps 8;  
QY 38 TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKKR 97  
DB 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAAKQVDEYIKKKL-----QLDRRK 44  
QY 98 HTITVALVNLNINIKVEYLKIVESTSSQILMMESRSKVDKAEVSKFEDSSSSSSD 157  
DB 45 HTQNVGLLTGLGVIKTEYLHGLSVSKKSEAE-LPSEIKAKLDAAPFQFKDT----- 96  
QY 158 SSTKPEASDTAKPNKPTPEGEKVAEAKKVEAEAKKADOKEDRRNYPNTYKTLLEI 217  
DB 97 -----LPTPEGKKVAEAEKKVEAEAKKADOKEDRRNYPNTYKTLLEI 142  
QY 218 AESDVEVKAELELVKVKANEPDRQKIQAEEVSKAEATRLKIKTDREAEAEAEAK 277  
DB 143 AESDVEVKAELELVKVEAKESRDEKKNQAKAVENKKAATRLKIKTDREKA-EEAK 201  
QY 278 RRADAK-----EQGPKGRAKGVPGELATPDKKENDAKSSDSSVGBETLPSPLK 328  
DB 202 RRADAKQEAENVATSEQDKSKRAKRVFGLATPDKKENDAKSSDSSVGBETLPSPLK 261  
QY 329 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 388  
DB 262 PEKKVAEAEKKVVEAKKAEADOKEDRRNYPNTYKTLLETAESDVEVKAELELVKEE 321  
QY 389 AKEPRNEEKVQKAEVSKAEATRLKIKTDKKA-EEAAKRAAEEDVKKEKPAQP 447  
DB 322 AKESRNEEKIKQVKAKVESKAEATRLKIKTDKKAEEAAKRAAEEDVKKEKPAQP 381  
QY 448 QPAPAPKAEK-----PAPAPKPNPAPQPAEKPADQAAE 481  
DB 382 QPAPAPQEPKTEEPENPAPAPKPNPAPKPAEKPADQAAE 424

## RESULT 10

US-09-286-981B-7  
; Sequence 7, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; CURRENT FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38



```
Db 225 DFGPATPKDNDKSSSSVGEETLPSLPGKGVAAEKKVVEEAKKAAQKQEDSH 284
Qy 356 RNYPTNTYKTLLEIAESDVEVKAALELVKEEAKPRNEEKVKQAKAEVSKAEATRL 415
Db 285 RNYPTITYKTLLEIAESDVEVKAALELVKEEAKGRNEEKVKQAKAEVSKAEATRL 344
Qy 416 EKIKTRDKAEAEAKRAAEEDVKVEKPAEQPAPAPAKAEKPAPAPKPPENPAEQPKAEK 475
Db 345 EKIKTRDKAEAEAKRAAEEDVKVEKPAEQPAPAPAKAEKPAPAPKPPENPAEQPKAEK 404
Qy 476 PADQQAEE 483
Db 405 PADQQAEE 412

RESULT 13
US-09-286-981B-16
; Sequence 16, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-16

Query Match 55.7%; Score 1417; DB 4; Length 414;
Best Local Similarity 67.7%; Pred. No. 2e-90;
Matches 310; Conservative 46; Mismatches 46; Indels 56; Gaps 10;

Qy 38 TENEGATQVPTSSNRANESQAEQGPCKLDSERDKARKE-VEEVVKIVGESYAKSTYK 96
Db 1 TENEGSTQAATSSNNAK-----TEHRKAAKQVVDEYIEKMLRE--IQLDRR 44
Qy 97 RHTITVALVNLNINKNEYLNK-IVESTSESQILMMESRSKYDVAVSKFEKDDSSSS 154
Db 45 RHTQVNLNINKUSAIKTYRLNVLNLEKSKDELPG---SEIKAKLDA--FEK----- 92
Qy 155 SSDSSTKPEASDTAKPNKPTPEGKVAEAKKVEEAKKADQKQEDRRNPTITYKTLE 214
Db 93 -----EKKDTLK-----PGKVAEAKKVEEAKKADQKQEDRRNPTITYKTLE 138
Qy 215 LEIASDVEVKAALELVKVKANEPDEQIKQAEAEVSKAEATRLKKIKTDREAE 274
Db 139 LEIADFVKVKEAELELVKEEAKESRNEGCTIKQAEKVESKAEATRLNIKTDRAE 198
Qy 275 EAKRADAK-----EQGPKGRAGVGPGLATPDKKENDAKSSDSSVGEETLPS 325
Db 199 EAKRADAKLEKNAVATSDQGPGRAGVGPGLATPDKKENDAKSSDSSVGEETLPS 258
Qy 326 SLKPKKVAEAEKKVVEEAKKAEADQKQEDRRNPTITYKTLEIAESDVEVKAALELV 385
Db 259 SLKSKGVAAEAEKKVVEEAKKADQKQEDRRNPTITYKTLLDLAEASDVKVEAELELV 318
Qy 386 KEEAKPRNEEKVKQAKAEVSKAEATRLKIKTDRAKAEAEAKRAAEEDVKVEKPAE 445
Db 319 KEEAKPRDEEKIKQAKAKVESKAEATRLNIKTDRAKAEAEAKRAAEEDVKVEKPAE 378
Qy 446 QOPAPAPAKAEKPAPAPKPPENPAEQPKAEKPADOQAE 483
Db 379 QOPAPAPQEPK--PAPKPEKPAEQPKAEKTDQQAEE 414
```

```
RESULT 14
US-09-286-981B-15
; Sequence 15, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-15

Query Match 55.6%; Score 1416; DB 4; Length 419;
Best Local Similarity 66.7%; Pred. No. 2.3e-90;
Matches 309; Conservative 43; Mismatches 49; Indels 62; Gaps 9;

Qy 38 TENEGATQVPTSSNRANESQAEQGPCKLDSERDKARKEVEEVVKIVGESYAKSTYK 97
Db 1 TENERTQVPTSSNRGK-----PERKAAEQDFEYINKMI-----OLDKRR 41
Qy 98 HTITVALVNLNINKNEYLNKIVESTSESQILMMESRSKYDVAVSKFEKDDSSSSSD 157
Db 42 HTQNLAFNQLSRIKTEYLNGLKEK-SEALP---SKIKASLDAAFKQFKDT----- 90
Qy 158 SSTPEASDTAKPNKPTPEGKVAEAKKVEEAE-----KKAQDQKQEDRRNPTITY 210
Db 91 -----LPTPEKKVAAEAEKKVVEEAKKVAEAKKAAQKQEDRRNPTITY 136
Qy 211 KTLLEIAESDVEVKAALELVKVKANEPDEQIKQAEAEVSKAEATRLKKIKTDRE 270
Db 137 KTLLEIAEFVKVKEAELELVKKEADSRNEGCTINQAKAKVESEKAEATRLKKIKTDRE 196
Qy 271 EA-BEEAKRRADAKEQG---KPKGRAGVGPGLATPDKKENDAKSSDSSVGEETLPS 326
Db 197 KAESEAKRRADAKEQDESRRKSRGKRGALGEQATPDKKENDAKSSDSSVGEETLPS 256
Qy 327 LKPEKVAEAEKKVVEEAKKAEADQKQEDRRNPTITYKTLEIAESDVEVKAALELVK 386
Db 257 LKPGKVAEAEKKVVEEADKKAKAQKQEDRRNPTITYKTLEIAESDVKVEAELELVK 316
Qy 387 EEAKEPRNEEKVKQAKAEVSKAEATRLKIKTDRAKAEAEAKRAAEEDVKVEKPAE 446
Db 317 EEAKEPRNEEKVKQAKAEVSKAEATRLKIKTDRAKAEAEAKRAAEEDVKVEKPAE 376
Qy 447 QOPAPAPAKAEKP-----PAPKPPENPAEQPKAEKPADOQAE 482
Db 377 QOPAPAPQEPKPAEPEPNVPAPKPPENPAEQPKAEKPADOQAE 419

RESULT 15
US-09-286-981B-10
; Sequence 10, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
```

```
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-10

Query Match      55.5%; Score 1413; DB 4; Length 414;
Best Local Similarity 66.6%; Pred. No. 3.7e-90;
Matches 305; Conservative 49; Mismatches 48; Indels 56; Gaps 8;

Qy 38 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKE-VVEYVVKIVGESYAKSTKK 96
Db 1 TENEGTQAATSSNMAK-----TEHRKAQVVDVEYIEKMLE--IQLDRR 44

Qy 97 RHTITVALVNLNINIKNEYLNK--IVESTSESQILMMESRSKVDEAVSFEKDSSSSS 154
Db 45 KHTQNVALNIIKLSAIIKTKYLRLNVLEEKSKDELP---SEIKAKLDAAFEKFKDT---- 97

Qy 155 SSDSSTKEASDTAKPNKPTPEGKVAEAKKKVEAEKKAKDOKEEDRRNYPTITYKTLE 214
Db 98 -----LKPGEKVAEAKKKVBEAKKAEQKEEDRRNYPTITYKTLE 138

Qy 215 LEIAESDVEVKAELELVKVKANPRDEQIKQAEAEVESQAEATRLKKIKTDREBAEE 274
Db 139 LEIAEFDVKVKEAELELVKEAEKSRNEGTIQAEKVESKAEATRLNIKTDRKAAEE 198

Qy 275 EAKRRADAK-----EQGPKGRAGVPGELATPDKKENDAKSDSSVGEETLPSP 325
Db 199 EAKRKADAKLEANVATSDQGPKGRAGVPGELATPDKKENDAKSDSSVGEETLPSS 258

Qy 326 SLKPEKKVAEAEKKVVEAEKKAEDQKEEDRRNYPTITYKTLEIAESDVEVKAELELV 385
Db 259 SLKSGKKVAEAEKKVVEAEKKAEDQKEEDRRNYPTITYKTLDLEIAESDVVKAELELV 318

Qy 386 KEEAKEPRNEEKVKQAEAEVESKAEATRLKEIKTDRKAAEEAEKRAAEEDKVKEKPAE 445
Db 319 KEEAKEPRDEEKIKQAEAEVESKAEATRLNIKTDRDAEEAEKRAAEEDKVKEKPAE 378

Qy 446 QPQAPAPKAEKPAPKPPENPAPQPKAEKPADQQAEE 483
Db 379 QPQAPAPATQPEK--PAPKPEKPAEQPKAEKTTDQQAEE 414
```

Search completed: September 28, 2005, 18:44:16  
Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2005, 18:32:15 ; Search time 178 Seconds  
(without alignments)  
1475.824 Million cell updates/sec

Title: US-09-298-523D-13

Perfect score: 2546

Sequence: 1 MFASKSRKVKHYIRKFSVG.....NRLTQQPPKTEKPAQPSTP 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_prot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2546	100.0	693	2	Q97N74	Q97N74 streptococc
2	2337	99.6	693	2	Q9K21	Q9K21 streptococc
3	2503.5	98.3	696	2	Q9KK32	Q9KK32 streptococc
4	2489.5	97.8	516	2	Q9RQT2	Q9RQT2 streptococc
5	2420	95.1	695	2	Q9KK36	Q9KK36 streptococc
6	2070.5	81.3	684	2	Q9KK17	Q9KK17 streptococc
7	1814	71.2	695	2	Q9KK51	Q9KK51 streptococc
8	1808	71.0	681	2	Q9KK16	Q9KK16 streptococc
9	1806.5	71.0	660	2	Q9KK18	Q9KK18 streptococc
10	1803	70.8	709	2	Q9KK38	Q9KK38 streptococc
11	1796.5	70.6	487	2	Q9RQT3	Q9RQT3 streptococc
12	1780.5	69.9	752	2	Q9KK30	Q9KK30 streptococc
13	1762.5	69.2	752	2	Q9KK33	Q9KK33 streptococc
14	1756	69.0	564	2	Q69188	Q69188 streptococc
15	1756	69.0	701	2	Q9KK48	Q9KK48 streptococc
16	1756	69.0	701	2	Q8DN05	Q8DN05 streptococc
17	1752	68.8	701	2	Q9RQT5	Q9RQT5 streptococc
18	1751.5	68.8	730	2	Q9KK47	Q9KK47 streptococc
19	1744	68.5	581	2	Q33742	Q33742 streptococc
20	1741.5	68.4	660	2	Q9KK41	Q9KK41 streptococc
21	1739.5	68.3	680	2	Q9KK43	Q9KK43 streptococc
22	1739.5	68.3	730	2	Q9KK31	Q9KK31 streptococc
23	1719	67.5	699	2	Q9KK42	Q9KK42 streptococc
24	1709	67.1	655	2	Q9KK50	Q9KK50 streptococc
25	1661	65.2	657	2	Q9KK29	Q9KK29 streptococc
26	1642	64.5	684	2	Q9KK46	Q9KK46 streptococc
27	1635.5	64.2	488	2	Q9RQT6	Q9RQT6 streptococc
28	1621.5	63.7	678	2	Q9KK54	Q9KK54 streptococc
29	1619.5	63.6	678	2	Q9KK49	Q9KK49 streptococc
30	1612.5	63.3	680	2	Q9KK11	Q9KK11 streptococc
31	1608	63.2	681	2	Q9KK13	Q9KK13 streptococc

32	1599	62.8	676	2	Q9KK39	Q9KK39 streptococc
33	1599	62.8	696	2	Q9KK14	Q9KK14 streptococc
34	1581	62.1	663	2	Q30874	Q30874 streptococc
35	1418	55.7	739	2	Q9RQT4	Q9RQT4 streptococc
36	1371	53.8	820	2	Q9RQT1	Q9RQT1 streptococc
37	1363.5	53.6	869	2	Q9KK27	Q9KK27 streptococc
38	1356	53.3	929	2	Q9KK19	Q9KK19 streptococc
39	1356	53.3	929	2	Q9ZAY5	Q9ZAY5 streptococc
40	1239	48.7	565	2	Q9KK15	Q9KK15 streptococc
41	1228.5	48.3	769	2	Q9KK40	Q9KK40 streptococc
42	1215	47.7	667	2	Q9KK28	Q9KK28 streptococc
43	1195.5	47.0	770	2	Q9KK37	Q9KK37 streptococc
44	1191.5	46.8	769	2	Q9FDQ1	Q9FDQ1 streptococc
45	1189	46.7	523	2	Q33753	Q33753 streptococc

#### ALIGNMENTS

RESULT 1  
Q97N74 ID Q97N74 PRELIMINARY; PRT; 693 AA.  
AC Q97N74;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Choline binding protein A.  
GN OrderedLocusNames=SP2190;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC BAA-334 / TIGR4;  
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Uterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,  
RA Hickey E.K., Holt I.E., Loftus B.J., Yang E., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."  
RL Science 293:498-506(2001).  
DR EMBL; AE007507; AAK76241.1; -.  
DR PIR; H95255; H95255.  
DR HSSP; P06653; LH86.  
DR TIGR; SP2190; -.  
GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; GPOG\_YSIK.  
DR InterPro; IPR007756; RICH.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF04650; YSIK signal; 1.  
DR TIGRFAMs; TIGR01168; YSIK\_signal; 1.  
KW Complete proteome.  
SQ SEQUENCE 693 AA; 77762 MW; 6F8F47C32E344A41 CRC64;

Query Match 100.0%; Score 2546; DB 2; Length 693;

Best Local Similarity 100.0%; Pred. No. 2.3e-80;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFASKSRKVKHYIRKFSVGVSVVVATENEGATQVPTSSNRANESQAEQ 60

Db 1 MFASKSRKVKHYIRKFSVGVSVVVATENEGATQVPTSSNRANESQAEQ 60

Qy 61 GEQPKLDSERDKARKEVEEVKKIVGESYAKSKTKRHTITVALVNEUNNKIYLNKIV 120

Db 61 GEQPKLDSERDKARKEVEEVKKIVGESYAKSKTKRHTITVALVNEUNNKIYLNKIV 120

Qy	121	ESTSESQQLIMWGRSKYDEAVSFKEKDSSTSSSSSSTKPEASDTAKPNKPTEPGEKV	180
Dd	121	ESTSESQQLIMWGRSKYDEAVSFKEKDSSTSSSSSSTKPEASDTAKPNKPTEPGEKV	180
Qy	181	AEAKKKVEAAEKAKDKOEEDRRNPTTYTKTLELEIAESDVEVKKAELVLVVKANEP	240
Dd	181	AEAKKKVEAAEKAKDKOEEDRRNPTTYTKTLELEIAESDVEVKKAELVLVVKANEP	240
Qy	241	DEQIKIQAABVESKQAEATRLKKITDRBAAEAEKRRADAKEQGPKGAKRGVPGL	300
Dd	241	DEQIKIQAABVESKQAEATRLKKITDRBAAEAEKRRADAKEQGPKGAKRGVPGL	300
Qy	301	ATPDKKENDAKSDSSVGEEITLPSLSLKPEKKVAEAEEKVVEAKKKAEDQKEDRRNYP	360
Dd	301	ATPDKKENDAKSDSSVGEEITLPSLSLKPEKKVAEAEEKVVEAKKKAEDQKEDRRNYP	360
Qy	361	NTYKTLELEIAESDVEVKKAELVLVVKANEP	420
Dd	361	NTYKTLELEIAESDVEVKKAELVLVVKANEP	420
Qy	421	DRKKAEEEEAKRAAEDKVKKEPAEQOPAPAKPAEPAPAKPENPAEQPKAEKPADQQ	480
Dd	421	DRKKAEEEEAKRAAEDKVKKEPAEQOPAPAKPAEPAPAKPENPAEQPKAEKPADQQ	480
Qy	481	AEDYARRSEEEYNRLTQQPPKTEKPAQPSTP	513
Dd	481	AEDYARRSEEEYNRLTQQPPKTEKPAQPSTP	513
<b>PSP-PT 2</b>			
ID	O9KK21	PRELIMINARY; PRT; 693 AA.	
AC	O9KK21;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Surface protein PspC.		
GN	Name=pspC;		
OS	Streptococcus pneumoniae.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1313;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=g9;		
RX	MEDLINE=21988621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;		
RT	Iannelli F., Oggioni M.R., Pozzi G.;		
RA	"Allelic variation in the highly polymorphic locus pspC of		
RL	Streptococcus pneumoniae.";		
RT	Gene 284:63-71(2002)		
DR	ENBL; AF154035; AAF73807.1; -.		
DR	HSP; P06653; IH8G.		
DR	GO; GO:0016020; C.membrane; IEA.		
DR	InterPro; IPR002479; CW_binding.		
DR	InterPro; IPR005877; GPcs YSIK.		
DR	InterPro; IPR007756; RICH-		
DR	Pfam; PF01473; CW_binding_1; 8.		
DR	Pfam; PF05062; RICH; 1.		
DR	Pfam; PF04650; YSIK_signal; 1.		
DR	TIGRFAMS; TIGR01168; YSIK_signal; 1.		
SQ	SEQUENCE 693 AA; 77789 MW; 8DA1BA111565667 CRC64;		
Query Match 99.6%; Score 2537; DB 2; Length 693;			
Best Local Similarity 99.8%; Pred. No. 4.7e-80;			
Matches 512; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	MFASKERKHVSIRKFSGVASVVVASLVMSVVHATENEGATQVPTSSNRANESQAEQ	60
Dd	1	MFASKERKHVSIRKFSGVASVVVASLVMSVVHATENEGATQVPTSSNRANESQAEQ	60
Qy	61	GEOPKLDSDERDKARKEVEYKKIVGESYAKTKGRHTTTVALVNELNNIKNEYLNKTV	120

```

Db 61 GQPKKLDSEKARKEVEYVKIIGESYAKSTKQRHTITVALVNLNNIKNEYLNKIV 120
QY 121 ESTSESQILMMESRSKVDEAVSKFEKDSOSSSSSSSSSTKPEASDTAKPNKPTPEGKV 180
Db 121 ESTSESQILMMESRSKVDEAVSKFEKDSOSSSSSSSSSTKPEASDTAKPNKPTPEGKV 180
QY 181 AEAKKKVEAEAKKADQKEEDRRNPTTYTKTLELEIAESDVVVKAELELVKVKANEP 240
Db 181 AEAKKKVEAEAKKADQKEEDRRNPTTYTKTLELEIAESDVVVKAELELVKVKANEP 240
QY 241 DEOKIKQAEAVESKQAEATRLKIKTDREAEAEAKRADAKEOGKPKRAKRGV 296
Db 241 DEOKIKQAEAVESKQAEATRLKIKTDREAEAEAKRADAKEOGKPKRAKRGV 300
QY 297 PGELATPDCKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVBEAKKABDQKEEDRR 356
Db 301 PGEQATPDCKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVBEAKKABDQKEEDRR 360
QY 357 NYPTTYTKTLELEIAESDVVVKAELELVKBEAKPRNEEKVKQAEVSKKAEATRLK 416
Db 361 NYPTTYTKTLELEIAESDVVVKAELELVKBEAKPRNEEKVKQAEVSKKAEATRLK 420
QY 417 KIKTDKKAEEAKKAAEEDVKKEKPAEQOPAPAPKAEKPAKAPKPNPAEQPIAEKP 476
Db 421 KIKTDKKAEEAKKAAEEDVKKEKPAEQOPAPAPKAEKPAKAPKPNPAEQPIAEKP 480
QY 477 ADQQAEDYARSEEEYNRLTQQQPPKTEKPAQSTP 513
Db 481 -DQQAEDYARSEEEYNRLTQQQPPKTEKPAQSTP 516

```

## RESULT 4

```

Q9ROT2 PRELIMINARY; PRT; 516 AA.
ID Q9ROT2 AC Q9K36 PRELIMINARY; PRT; 516 AA.
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein pspC (Fragment).
GN Name=pspC;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L81905.
RX MEDLINE=20038119; PubMed=10569772;
RA Brooks-Walter A., Briles D.B., Hollingshead S.K.;
RT "The pspC gene of Streptococcus pneumoniae encodes a polymorphic
RT protein, PspC, which elicits cross-reactive antibodies to PspA and
RT provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67, 6533-6542 (1999).
DR EMBL; AF068649; AAF13459.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR007756; RICH.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
KW Hypothetical protein.
FT NON_TER 516
SQ SEQUENCE 516 AA; 57886 MW; 1119782688BB7E35 CRC64;

```

```

Query Match 97.8%; Score 2489.5; DB 2; Length 516;
Best Local Similarity 98.2%; Pred. No. 1.6e-78;
Matches 504; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 1 MPASKSERKVHYSIRKFSVGVSFVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MPASKSERKVHYSIRKFSVGVSFVVASLVMSGVVHATENEGLPST-TSSNRANESQAEQ 59

```

```

QY 61 GQPKKLDSEKARKEVEYVKIIGESYAKSTKQRHTITVALVNLNNIKNEYLNKIV 120
Db 60 GQPKKLDSEKARKEVEYVKIIGESYAKSTKQRHTITVALVNLNNIKNEYLNKIV 119
QY 121 ESTSESQILMMESRSKVDEAVSKFEKDSOSSSSSSSSSTKPEASDTAKPNKPTPEGKV 180
Db 120 ESTSESQILMMESRSKVDEAVSKFEKDSOSSSSSSSSSTKPEASDTAKPNKPTPEGKV 179
QY 181 AEAKKKVEAEAKKADQKEEDRRNPTTYTKTLELEIAESDVVVKAELELVKVKANEP 240
Db 180 AEAKKKVEAEAKKADQKEEDRRNPTTYTKTLELEIAESDVVVKAELELVKVKANEP 239
QY 241 DEOKIKQAEAVESKQAEATRLKIKTDREAEAEAKRADAKEOGKPKRAKRGV 300
Db 240 DEOKIKQAEAVESKQAEATRLKIKTDREAEAEAKRADAKEOGKPKRAKRGV 299
QY 301 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVBEAKKABDQKEEDRRNPT 360
Db 300 ATPDKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVBEAKKABDQKEEDRRNPT 359
QY 361 NYTKTLELEIAESDVVVKAELELVKBEAKPRNEEKVKQAEVSKKAEATRLK 420
Db 360 NYTKTLELEIAESDVVVKAELELVKBEAKPRNEEKVKQAEVSKKAEATRLK 419
QY 421 DRKAAEEAKKAAEEDVKKEKPAEQOPAPAPKAEKPAKAPKPNPAEQPIAEKP 480
Db 420 DRKAAEEAKKAAEEDVKKEKPAEQOPAPAPKAEKPAKAPKPNPAEQPIAEKP 479
QY 481 AEEDYARSEEEYNRLTQQQPPKTEKPAQSTP 513
Db 480 AEEDYARSEEEYNRLTQQQPPKTEKPAQSTP 512

```

## RESULT 5

```

Q9KK36 PRELIMINARY; PRT; 695 AA.
ID Q9KK36 AC Q9K36;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Surface protein pspC.
GN Name=pspC;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G394;
RX MEDLINE=21888621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic variation in the highly polymorphic locus pspC of
RT Streptococcus pneumoniae.";
RL Gene 284, 63-71 (2002).
DR EMBL; AF154023; AAF73792.1; -.
DR HSSP; P06653; IH8G.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR007756; RICH.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 695 AA; 78361 MW; F8C84F08C4EFFFBA CRC64;

```

```

Query Match 95.1%; Score 2420; DB 2; Length 695;
Best Local Similarity 95.6%; Pred. No. 4.9e-76;
Matches 494; Conservative 6; Mismatches 11; Indels 6; Gaps 2;

QY 1 MPASKSERKVHYSIRKFSVGVSFVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MPASKSERKVHYSIRKFSVGVSFVVASLVMSGVVHATEKATQVPTSSNRANESQAEQ 60

```





QY 1 MPASKSERKVHYSIRKFSVGVASVAVSLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
 DB 1 MPASKSERKVHYSIRKFSVGVASVAVSLVMSGVVHATENEGSTQAATSSNNAK 54  
 QY 61 GEQPKCLDSEDRKARKEVEEVVKITVGSYAKSTKKRHTITVALVNLNNTKNEVYNKIV 120  
 DB 55 -----TEHMKAAQVDEYIKKKL-----QLDRKHTQNVGLTKLVKITEYLHGLS 101  
 QY 121 ESTSESQILMMESRSKVDEAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGSKV 180  
 DB 102 VSKESEAE-LPSEITAKLDAAFQPKKOT-----LPTPEGKKV 139  
 QY 181 AEAKKVEAEAKKADQKEEDRRNYPTITYKTLELEIAESDVEVKAELELVKVKANEP 240  
 DB 140 AEAEKKVEAEAKKADQKEEDRRNYPTITYKTLELEIAESDVEVKAELELVKKEAKESR 199  
 QY 241 DEQIKKQAEAEVESKQAEATRLKKITDREAEAEAKRADAKQEG---KPKGRKRGVP 297  
 DB 200 DEKKINQAKAVESQAEATRLKKITDREAEAEAKRADAKQEGDESKRRKSRVKGDL 259  
 QY 298 GELATPDKKENDAKSSDSVGBETLPSLSLKPEKKVAEAEKKVEAEAKKADQKEEDRRN 357  
 DB 260 GEPATPDKKENDAKSSDSVGBETLPSLSLKPGKKVAEAEKKVEAEAKKADQKEEDHRN 319  
 QY 358 YPTNTYKTLELEIAESDVEVKAELELVKKEAEKPRNEEKVKQAKAEVESKAEATRLK 417  
 DB 320 YPTITYKTLELEIAESDVEVKAELELVKKEAKGSRNEEKVKQAKAEVESKAEATRLK 379  
 QY 418 IKTDKKAEEAEAKRAAEEDVKKEPAQOPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 477  
 DB 380 IKTDKKAEEAEAKRAAEEDVKKEPAQOPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 439  
 QY 478 DQAEEDYARSEEEYNRLTQQPPKTEKPAQPS 513  
 DB 440 DQAEEDYARSEEEYNRLTQQPPKTEKPAQPS 475

## RESULT 8

Q9KK16 PRELIMINARY; PRT; 681 AA.  
 AC Q9KK16;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Surface protein PcpC.  
 GN Name=pspC;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=serf25;  
 RX MEDLINE=21888621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic variation in the highly polymorphic locus pcpC of  
 ST Streptococcus pneumoniae.";  
 RL Gene 284:63-71(2002).  
 DR EMBL; AF154040; AAF73812.1; -.  
 DR HSSP; P06653; 1HCG.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR005877; Gpos YSIRK.  
 DR InterPro; IPR007756; RICH-  
 DR Pfam; PF01473; CW binding\_1; 8.  
 DR Pfam; PF05062; RICH; 1.  
 DR Pfam; PF04650; YSIRK signal; 1.  
 DR TIGRFAMs; TIGR01168; YSIRK signal; 1.  
 SQ SEQUENCE 681 AA; 76728 MW; 38FE4782653D51A8 CRC64;

Query Match 71.0%; Score 1808; DB 2; Length 681;  
 Best Local Similarity 70.8%; Pred. No. 5e-55;

Matches 380; Conservative 44; Mismatches 51; Indels 62; Gaps 7;  
 QY 1 MPASKSERKVHYSIRKFSVGVASVAVSLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
 DB 1 MPASKSERKVHYSIRKFSVGVASVAVSLVMSGVVHATEKEVTTQVPTSSNRANKSQ--- 57  
 QY 61 GEQPKCLDSEDRKARKEVEEVVKITVGSYAKSTKKRHTITVALVNLNNTKNEVYNKIV 120  
 DB 58 -----TEHMKAAQVDEYIKKKL-----QLDRKHTQNVGLTKLVKITEYLHGLS 104  
 QY 121 ESTSESQILMMESRSKVDEAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGSKV 180  
 DB 105 VSKESEAE-LPSEVAKLDAAFQPKKOT-----LPTPEGKKV 142  
 QY 181 AEAKKVEAEAKKADQKEEDRRNYPTITYKTLELEIAESDVEVKAELELVKVKANEP 240  
 DB 143 AEAEKKVEAEAKKADQKEEDRRNYPTITYKTLELEIAESDVEVKAELELVKGGYKPR 202  
 QY 241 DEQIKKQAEAEVESKQAEATRLKKITDREAEAEAKRADAK-----EQGKP 288  
 DB 203 DEQVNOAKAVESKQAEATRLKKITDREAEAEAKRADAKKEAVEKNTATSEQKP 262  
 QY 289 KGRKRGVPGELATPDKKENDAKSSDSVGBETLPSLSLKPEKKVAEAEKKVEAEAKKAE 348  
 DB 263 KRGKRGALGEOATPDKKENDAKSSDSVGBETLPSLSLKPEKKVAEAEKKVEAEAKKAK 322  
 QY 349 DQKEEDRRNYPTNTYKTLELEIAESDVEVKAELELVKKEAEKPRNEEKVKQAKAEVESK 408  
 DB 323 DQKEEDRRNYPTNTYKTLELEIAESDVEVKAELELVKKEAEKSRNEEKIKQVRAKVESK 382  
 QY 409 KAEATRLKIKTDKKAEEAEAKRAAEEDVKKEPAQOPAPAPAPAPAPAPAPAPAPAPAP 461  
 DB 383 KAEATRLKIKTDKKAEEAEAKRAAEEDVKKEPAQOPAPAPAPAPAPAPAPAPAPAPAP 442  
 QY 462 PKPENPAQPKAEKPADQAEEDYARSEEEYNRLTQQPPKTEKPA-----QPSTP 513  
 DB 443 PKPENPAQPKAEKPADQAEEDYARSEEEYNRLTQQPPKTEKPAPEPAPPEPAP 499  
 RESULT 9  
 Q9KK18 PRELIMINARY; PRT; 660 AA.  
 AC Q9KK18;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Surface protein PcpC.  
 GN Name=pspC;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=serf2;  
 RX MEDLINE=21888621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic variation in the highly polymorphic locus pcpC of  
 ST Streptococcus pneumoniae.";  
 RL Gene 284:63-71(2002).  
 DR EMBL; AF154038; AAF73810.1; -.  
 DR HSSP; P06653; 1H8G.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR005877; Gpos YSIRK.  
 DR InterPro; IPR007756; RICH-  
 DR Pfam; PF01473; CW binding\_1; 8.  
 DR Pfam; PF05062; RICH; 1.  
 DR Pfam; PF04650; YSIRK signal; 1.  
 DR TIGRFAMs; TIGR01168; YSIRK signal; 1.  
 SQ SEQUENCE 660 AA; 74515 MW; 374EB582FD09A659 CRC64;

Query Match

71.0%; Score 1806.5; DB 2; Length 660;

```
Best Local Similarity 73.6%; Pred. No. 5.5e-55;
Matches 384; Conservative 39; Mismatches 48; Indels 51; Gaps 7;

Qy 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMSVHVHATENEGATQVPTSSNRANESQAQ 60
Db 1 MFASKNERKVVHSIRKFSIGVASVAVASLPMGVSVVHATKRVTTQVATSSNKANKSQ--- 57
Qy 61 GEQPKKLDSEKDKARKEVEEYVKIIVGESYAKSTKKRHTITVALVNLNNIKN 120
Db 58 -----TEYKAAKQVDEYITKL-----QLDRREHIQNVGLLTKLVIKTEYLHRLS 104
Qy 121 ESTSESQIILMMESRSKVDVAESKPKDSSSSSSSSSTKPEASDTAKPNKPTPEGKV 180
Db 105 VSKEKSEAE-LPSIKAKLDAAFELEFKDT-----LPTPEGKV 142
Qy 181 AEAKKVEEAEKKAQDOKEEDRRNYPTTYKTLELEIAESDVEVKKAELELVKANEPR 240
Db 143 AEAEKVEEAEKKAQDOKEEDRRNYPTTYKTLELEIAESDVEVKKAELELVKEEAKESR 202
Qy 241 DEQIKQAEAEVESKQAEATRLKKIKTDREAEAEAKRRADAK-----EQGPKGR 291
Db 203 DEKKINQAKAVENKKAETELKNIKTDREKA-EAEKRRADAKLQEANVATSEODKSKRR 261
Qy 292 AKRGVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKVAEAEKVKKAAEDQK 351
Db 262 ANREVLGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKVAEAEKVKKAAEDQK 321
Qy 352 EEDRRNYPTTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVESKAE 411
Db 322 EEDRRNYPTTYKTLELEIAESDVEVKKAELELVKEEAKPRNEEKVKQAKAEVESKAE 381
Qy 412 ATRLEKIKTDRAKAEAEAKRAAEEDKVKEPAEQOPAPAPKAEKPPAPKPNPAEQ 471
Db 382 ATRLEKIKTDRAKAEAEAKRAAEEDKVKEPAEQOPAPAPKAEKPPAPKPNPAEQ 438
Qy 472 KAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSTP 513
Db 439 KAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSTP 480

RESULT 10
Q9KK38 PRELIMINARY; PRT; 709 AA.
AC Q9KK38;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Surface protein PspC.
GN Name=pspC;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g387;
RX MEDLINE=21888621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic variation in the highly polymorphic locus pspC of
Streptococcus pneumoniae.";
RL Gene 284:63-71(2002).
DR EMBL; AF154021; AAF73790.1; -.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR007756; RICH.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF01473; Tropomyosin.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR PRINTS; PR00194; TROPOMYSIN.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
Hypothe
```

```
SQ SEQUENCE 709 AA; 80251 MW; 32BBC96E380EBB7A CRC64;
Query Match 70.8%; Score 1803; DB 2; Length 709;
Best Local Similarity 71.1%; Pred. No. 7.6e-55;
Matches 382; Conservative 49; Mismatches 54; Indels 52; Gaps 9;

Qy 1 MFASKSERKVVHSIRKFSVGVASVVVASLVMSVHVHAT-ENEGATQVPTSSNRANESQ-- 57
Db 1 MFASKNERKVVHSIRKFSIGVASVAVASLPMGVSVVHATGCGNYPSSPCSSSEANKSQTE 60
Qy 58 -----AEQGEQPKLDSEKDKARKEVEEYVKIIVGESYAKSTKKRHTITVALVNLNNIKN 113
Db 61 HMKAAEQGQKRRHIDLRNKAIDYEIKMLSE--IQDKRKHTQNFALNKLKSLRIKT 118
Qy 114 EYLN--KIVSTESOLOI---LMMESRSKVDVAESKPKDSSSSSSSSSTKPEASDTA 168
Db 119 EYLYGLSVLKESEAELEPSKAELEPSKIKELTAAPFEHFKDT----- 160
Qy 169 KPNKPTPEGKVAEAEKVKVEAEKKAQDOKEEDRRNYPTTYKTLELEIAESDVEVKKAE 228
Db 161 -----LRPGEKVAEAEKVKVEAEKKAQDOKEEDRRNYPTTYKTLELEIAESDVEVKKAE 215
Qy 229 LELVKVANEPRDEOKIKQAEAEVESKQAEATRLKKIKTDREAEAEAKRADAK----- 283
Db 216 LELVKEANPRNEQVKQAKAVESKQAEATRLKKIKTDREAEAEAKRADAKKEAV 275
Qy 284 -----EQGPKGRAGRVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKVAEA 336
Db 276 EKMTATSEQGPKGRAGRGALGEPATPDKKENDAKSSDSVSGEETLPSPSLKPEKVAEA 335
Qy 337 EKKVEAEAKKAQDOKEEDRRNYPTTYKTLELEIAESDVEVKKAELELVKEEAKPRNEE 396
Db 336 EKKVEAEAKKAQDOKEEDRRNYPTTYKTLDLSAESDVKVEAELELVKEEAKPRDEE 395
Qy 397 KVVQAKAEVESKQAEATRLKKIKTDRAKAEAEAKRAAEEDKVKEPAEQOPAPAPKAE 456
Db 396 KIKQAEKVESKQAEATRLKKIKTDRAKAEAEAKRAAEEDKVKEPAEQOPAPAPQPE 454
Qy 457 KPAPAPKPNPAEQKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSTP 513
Db 455 K--PAPKPEKPAEHPKAEKPADQQAEDYARRSEEEYNRLTQQPPKTEKPAQPSTP 509

RESULT 11
Q9RQT3 PRELIMINARY; PRT; 487 AA.
AC Q9RQT3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein pspC (Fragment).
GN Name=pspC;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BG8090;
RX MEDLINE=20038319; PubMed=10569772;
RA Brooks-Walter A., Briles D.E., Hollingshead S.K.;
RT "The pspC gene of Streptococcus pneumoniae encodes a polymorphic
protein, pspC, which elicits cross-reactive antibodies to PspA and
provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67:6533-6542(1999).
DR EMBL; AF068648; AAF13458.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR007756; RICH.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
Hypothe
```



DR Pfam; PF05062; RICH; 1.  
DR Pfam; PF04650; YSIRK signal; 1.  
DR TIGRFAMs; TIGR01168; YSIRK signal; 1.  
SQ SEQUENCE 752 AA; 85321 MW; E24BPF1DBF50A7F8 CRC64;

Query Match 69.2%; Score 1762.5; DB 2; Length 752;  
Best Local Similarity 71.5%; Pred. No. 2e-53;  
Matches 368; Conservative 55; Mismatches 47; Indels 45; Gaps 6;

Qy 1 MFASKSRKVVHYSIRKFSVGASVVVSLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSRKVVHYSIRKFSVGASVVVSLVMSGVVHATENEGATQVPTSSNRANESQAEQ 57  
Qy 61 GEQPKLDSERDKARKEVEEYVKIVGESYAKTKGHTITVALVNLNINKYELNKKV 120  
Db 58 -----TEHMKAAEQVDEYVNMW-----QLDRKHTQNLALNKLKSAIKTKYLRELN 104  
Qy 121 ESTSESQILMMBSRSKVDPAVSKFEKDSSSSSSDSSTKPEASDTAKPNKPTPEGEKV 180  
Db 105 VLEKSKKEELTSKTKKEIDAAFEQFKDT-----LKPGEKV 141  
Qy 181 AEAKKVEEAKKADQKEEDRRNYPTITYKTLEIAESDVVEYKKALELVKVKANEPR 240  
Db 142 EEAQKVEEAKKADQKEEDRRNYPTITYKTLEIAESDVVEYKKALELVKVKAGSR 201  
Qy 241 DEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKRRADAEQ--GKPKGRAKRGVPG 298  
Db 202 NEEKIKKAAEVESKKAETKLEIKTERKAAEAEAKKAAEAEVVKDKLKKRTYRGALG 261  
Qy 299 ELATPDKKENDAKSSDSVSGEETLPSLSLKEPKKVAEAEKVEBAKKAEDQKEEDRRNY 358  
Db 262 EPATPDKKENDAKSSDSVSGEETLPSLSLKEPKKVAEAEKVAEAEKKAQDQKEEDRRNY 321  
Qy 359 PTNTYKTLEIAESDVVEYKKALELVKEAEKPRNEEKVKQAEEVESKKAETRLKKI 418  
Db 322 PTNTYKTLEIAESDVVEYKKALELVKEAEKPRNEEKVKQAEEVESKKAETRLKKI 381  
Qy 419 KTRDKKAAEAEAKKAAEEDVKKEPAEQOPAPAPAKPAEKPAKPAKPAEQPAEKPAD 478  
Db 382 KTRDKKA-EAAKRVAEEDVKKEPAEQOPAPAPAKPAEKPAEQPAEKPAD 437  
Qy 479 QAAEEDYARRSEEEYNRLTQQPPKTEKPAQPSPTP 513  
Db 438 QAAEEDYARRSEEEYNRLTQQPPKTEKPAQPSPTP 472

RESULT 14  
O69188  
ID O69188 PRELIMINARY; PRT; 564 AA.  
AC O69188;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE C3-binding protein.  
GN Name=pbca;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CP1200;  
RC MEDLINE=20281293; PubMed=10820017; DOI=10.1021/b1992157d;  
RA Cheng Q., Finkel D., Hostetter M.K.;  
RT "Novel purification scheme and functions for a C3-binding protein from  
Streptococcus pneumoniae.";  
RL Biochemistry 39:5450-5457(2000).  
DR EMBL; AF067128; AAC17445.1; -.  
DR HSSP; P06653; 1HCX.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpob\_Ysirk.  
DR InterPro; IPR007756; RICH.

DR Pfam; PF01473; CW binding\_1; 4.  
DR Pfam; PF05062; RICH; 1.  
DR Pfam; PF04650; YSIRK signal; 1.  
DR TIGRFAMs; TIGR01168; YSIRK signal; 1.  
SQ SEQUENCE 564 AA; 63596 MW; 884C97BAA0AED20 CRC64;

Query Match 69.0%; Score 1756; DB 2; Length 564;  
Best Local Similarity 70.9%; Pred. No. 2.6e-53;  
Matches 372; Conservative 50; Mismatches 47; Indels 56; Gaps 8;

Qy 1 MFASKSRKVVHYSIRKFSVGASVVVSLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSRKVVHYSIRKFSVGASVVVSLVMSGVVHATENEGSTQAATSSNMAK----- 54  
Qy 61 GEQPKLDSERDKARKEVEEYVKIVGESYAKTKGHTITVALVNLNINKYELNKK- 118  
Db 55 -----TEHMKAAEQVDEYIEKMLRE--IQDRKHTQNLALNKLKSAIKTKYLREL 104  
Qy 119 -IVETSESQILMMBSRSKVDPAVSKFEKDSSSSSSDSSTKPEASDTAKPNKPTPEG 177  
Db 105 NVLEKSKDELPT--SEIKAKLDAAPFKPKDT-----LKPGEKV 138  
Qy 178 EKVAAKVKVEEAEKKAQDQKEEDRRNYPTITYKTLEIAESDVVEYKKALELVKVKAN 237  
Db 139 EKVAAKVKVEEAEKKAQDQKEEDRRNYPTITYKTLEIAESDVVEYKKALELVKEAEK 198  
Qy 238 EPDEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKRRADAK-----EOKGP 288  
Db 199 ESRNEGTTIKAKKVESKKAETRLKKIKTDREAEAEAKRRADAKLKEANVATSDOGKP 259  
Qy 289 KGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPSLSLKEPKKVAEAEKVEEAEKKA 348  
Db 259 KGRAKRGVPGELATPDKKENDAKSSDSVSGEETLPSLSLKGKVKVAEAEKVEEAEKKA 318  
Qy 349 DQKEEDRRNYPTNTYKTLEIAESDVVEYKKALELVKEAEKPRNEEKVKQAEEVESK 408  
Db 319 DQKEEDRRNYPTNTYKTLDLEIAESDVVEYKKALELVKEAEKPRDEKIKQAQKAVESK 378  
Qy 409 KASATRLKKIKTDREAEAEKKAEEEDVKKEPAEQOPAPAPAKPAEKPAKPAKPAENPA 468  
Db 379 KASATRLKKIKTDREAEAEKKAEEEDVKKEPAEQOPAPAPAKPAEQPATQPEK--PAPKPEKPA 436  
Qy 469 EOPKAEKPADQAAEEDYARRSEEEYNRLTQQPPKTEKPAQPSPTP 513  
Db 437 EQPKAEKTDQAAEEDYARRSEEEYNRLTQQPPKTEKPAQPSPTP 481

RESULT 15  
O9KK48  
ID O9KK48 PRELIMINARY; PRT; 701 AA.  
AC O9KK48;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Surface protein PspC.  
GN Name=pspC;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=D39;  
RC MEDLINE=21888621; PubMed=11891047; DOI=10.1016/S0378-1119(01)00896-4;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic variation in the highly polymorphic locus pspC of  
Streptococcus pneumoniae.";  
RL Gene 284:63-71(2002).  
DR EMBL; AF154012; AAF73779.1; -.  
DR FIR; H98120; H98120.  
DR HSSP; P06653; 1H8G.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR002479; CW binding.

```
DR InterPro; IPR005877; Gpos_YsIRK.
DR InterPro; IPR007756; RICH.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF05062; RICH; 1.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 701 AA; 79098 MW; A80E31FE2846F1BF CRC64;

Query Match      69.0%; Score 1756; DB 2; Length 701;
Best Local Similarity 70.9%; Pred. No. 3.le-53;
Matches 372; Conservative 50; Mismatches 47; Indels 56; Gaps 8;

Qy 1 MFASKSERKRVHSYIRKFSVGVASVVGSLVVGSVVHATENEGATQVPTSSNRANESQAEQ 60
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 MFASKSERKRVHSYIRKFSVGVASVVGSLVVGSVVHATENEGSTQATSSNNAK----- 54
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 61 GEQPKLLDSERDKARKE-VEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNEYLNK- 118
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 55 -----TEHRAAKQVVDEYIEKMLRE--IQDDRKHQTQNALNINIKLSAIKTKYLREL 104
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 119 -IVSTSESQILMMESRSKVDEAVSKPEKSDSSSSSSSSSTKPEASDTAKPNKPTPG 177
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 105 NVLBEKSKDELFP--SEIKAKLDAAFKFKDT-----LKPQ 138
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 178 EKVAEAKKKVVEAEKKAKDKQEEEDRRNYPTTYKTLELEIAESDVEVKKAELELVKUKAN 237
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 139 EKVAEAKKKVVEAEKKAKDKQEEEDRRNYPTTYKTLELEIAEFQVVKKEAELELVKEEK 198
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 238 EPRDEQIKIQAEAVESQAEATRLKIKTDREAEAEAEKRRADAK-----EOGKP 288
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 199 ESRNEGTTIKQAEKVESKKAATRLNKTDRKAAEEAEKAKADAKLKEANVATSDQGKP 258
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 289 KGRAKGVPGELATPDKKENDAKSDSSVGBETLPSPSLKPEKKVAEAEKKVVEAEKKAE 348
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 259 KGRAKGVPGELATPDKKENDAKSDSSVGBETLPSSSLKSGKKVAEAEKKVVEAEKKAK 318
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 349 DQKEEDRRNYPTTYKTLELEIAESDVEVKKAELELVKEAEKEPRNEEKVKQAEVESK 408
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 319 DQKEEDRRNYPTTYKTLDLEIAESDVVKAELELVKEAEKEPRDEEKIKQAKAKVESK 378
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 409 KAEATRLKIKTDREAEAEAEKRAAEEDKVKEKPAEQOPAPAPAKPAKPENPA 468
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 379 KAEATRLNINIKTDREAEAEAEKRAAEEDKVKEKPAEQOPAPAPATQPEK--PAPKPKPA 436
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 469 EQPKAEKPADQQAEDYARRSEEEYNRLTQQOPPTEKPAQFSTP 513
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 437 EQPKAEKTDQQAEDYARRSEEEYNRLTQQOPPTEKPAQFSTP 481
Db |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: September 28, 2005, 18:42:39  
Job time : 181 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2005, 18:33:03 ; Search time 42 Seconds  
(without alignment)  
1175.219 Million cell updates/sec

Title: US-09-298-523D-13  
Perfect score: 2546  
Sequence: 1 MFASKSERKVVHYSIRKFSVG.....NRLTQQQPKTEKPAQPSTP 513

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: Pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2546	100.0	693	2 H95255	choline binding pr
2	1756	69.0	701	2 H98120	choline binding pr
3	496	19.5	619	2 A97887	surface protein ps
4	496	19.5	619	2 A41971	surface protein ps
5	397	15.6	744	2 F95013	pneumococcal surfa
6	393.5	15.5	1110	2 I51116	NF-180 - sea lampr
7	365.5	14.4	1164	1 FCSOAG	IgA Fc receptor pr
8	365	14.3	1020	1 QFHUH	neurofilament trip
9	361.5	14.2	1701	2 T09127	probable erythrocy
10	360.5	14.2	1134	2 A60234	IgA Fc receptor pr
11	343.5	13.5	606	2 A43427	neurofilament trip
12	338	13.3	1052	1 A44937	kinetoplast-associ
13	336.5	13.2	3488	2 T34418	hypothetical prote
14	334.5	13.1	6642	2 T29757	protein UNC-89 - C
15	329.5	12.9	1390	2 S51364	sperm tail-specifi
16	321	12.6	539	2 A28549	M24 protein precur
17	318.5	12.5	522	2 C96608	hypothetical prote
18	317	12.5	706	2 A45990	functional sarcop
19	314.5	12.4	1072	1 A37221	neurofilament trip
20	310.5	12.2	854	2 S02003	neurofilament trip
21	310	12.2	699	2 E84565	hypothetical prote
22	309.5	12.2	5327	2 T13564	microtubule-associ
23	300.5	11.8	501	2 A44643	M protein precursor
24	299.5	11.8	771	1 A33430	h-caldesmon - chic
25	297	11.7	1087	1 QFMNH	neurofilament trip
26	297	11.7	1815	2 C81169	IgA-specific metal
27	296.5	11.6	1094	2 S49313	protein kinase - s
28	294	11.5	858	2 S15762	neurofilament trip
29	292.5	11.5	1526	2 A45605	mature-parasite-in

30	291.5	11.4	729	2 S68191	triadin - human
31	290.5	11.4	1128	2 T30296	R27-2 protein - Tr
32	290	11.4	1773	2 A81937	IgA-specific metal
33	287	11.3	849	2 S00030	neurofilament trip
34	283	11.1	1192	2 A71623	probable secreted
35	281.5	11.1	644	2 S55395	neurofilament prot
36	281.5	11.1	916	2 A27864	neurofilament trip
37	280.5	11.0	587	2 JC1419	Fc gamma (IgG) rec
38	279.5	11.0	532	2 S54871	M protein - Strept
39	276.5	10.9	667	2 A40713	cylicin I - bovin
40	276	10.8	488	2 I46014	cylicin II - bovin
41	274.5	10.8	564	2 A60115	M protein precursor
42	273.5	10.7	2464	1 QRMSP1	microtubule-associ
43	273	10.7	332	2 B43427	neurofilament prot
44	270.5	10.6	598	2 B40713	cylicin I - human
45	270	10.6	483	2 A26297	M6 protein - Strep

ALIGNMENTS

RESULT 1

H95255  
choline binding protein A [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: H95255  
R:Tettelein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-693 <KUR>  
A:Cross-references: UNIPROT:Q9N74; GB:A6005672; PIDN:AAK76241.1; PID:g14973701; GSPDB:G1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP2190

Query Match 100.0%; Score 2546; DB 2; Length 693;  
Best Local Similarity 100.0%; Pred. No. 3.3e-89;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MFASKSERKVVHYSIRKFSVGASVVMGSLVMSVGHATENEGATQVPTSSNRANESQAEQ	60
Db	1	MFASKSERKVVHYSIRKFSVGASVVMGSLVMSVGHATENEGATQVPTSSNRANESQAEQ	60
Qy	61	GEQPKLDSERDKARKEVEEYVKKIVGESYAKSTKGRHTITVALVNLNLIKNEYLNKIV	120
Db	61	GEQPKLDSERDKARKEVEEYVKKIVGESYAKSTKGRHTITVALVNLNLIKNEYLNKIV	120
Qy	121	ESTSESQILIMESRSKVDVAVKFEKSSSSSSSSSSSTKPEASDTPAKNKPTEPGSKV	180
Db	121	ESTSESQILIMESRSKVDVAVKFEKSSSSSSSSSTKPEASDTPAKNKPTEPGSKV	180
Qy	181	AEAKKVEEAEEKKADQKEEDRRNYPITYKLTLEIAESDVEVKKAELVLVKVKANPR	240
Db	181	AEAKKVEEAEEKKADQKEEDRRNYPITYKLTLEIAESDVEVKKAELVLVKVKANPR	240
Qy	241	DEQIKQAEAEVESKQAEATRLKKIKTDREBAEAEKRRADAKGQKPKGRAGVPGEL	300
Db	241	DEQIKQAEAEVESKQAEATRLKKIKTDREBAEAEKRRADAKGQKPKGRAGVPGEL	300
Qy	301	ATPDKKENDAKSSSSVGEETLPSPSLUKPEKKVAEAKKVEEAKKADQKEEDRRNYP	360
Db	301	ATPDKKENDAKSSSSVGEETLPSPSLUKPEKKVAEAKKVEEAKKADQKEEDRRNYP	360
Qy	361	NTYKLTLEIAESDVEVKKAELVLVKVKANPRNEEKVKQAKAEVSKKAEATLEKTKT	420
Db	361	NTYKLTLEIAESDVEVKKAELVLVKVKANPRNEEKVKQAKAEVSKKAEATLEKTKT	420

Db 361 NTYKTLLEIAESDVVVKAELELVKEAEKPRNEEKVKQAKAEVSKAEATRLKLEIKT 420  
Qy 421 DRKKAEEAEKAAEEDKVKEPAEQPAPAPKAEKPAPAPKAPENPAEQPAEKAPADQ 480  
Db 421 DRKKAEEAEKAAEEDKVKEPAEQPAPAPKAEKPAPAPKAPENPAEQPAEKAPADQ 480  
Qy 481 AEEDYARRSEYENRLTQQPPKTEKPAQSTP 513  
Db 481 AEEDYARRSEYENRLTQQPPKTEKPAQSTP 513

RESULT 2  
H98120  
choline binding protein A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: H98120  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: H98120  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-701 <KUR>  
A:Cross-references: UNIPROT:Q9KK48; UNIPROT:Q8DN05; GB:AE007317; PIDN:AAL00797.1; PID:gl  
C:Genetics:  
A:Gene: pspC

Query Match 69.0%; Score 1756; DB 2; Length 701;  
Best Local Similarity 70.9%; Pred. No. 1.6e-59;  
Matches 372; Conservative 50; Mismatches 47; Indels 56; Gaps 8;

Qy 1 MFASKSERKHYISIRKFSVGVASVAVSLVMSVGHATENEGATQVPTSSNRANESQAE 60  
Db 1 MFASKSERKHYISIRKFSVGVASVAVSLVMSVGHATENEGSTOATSSNMAK----- 54  
Qy 61 GEQPKLDSERDKARKE-VEYVVKIVGESVAKTKKRRHTITVALVNELNINIKNEYLNK- 118  
Db 55 -----TEHRKAQKVVDYIEIKMRE--IQDRLRRKHTONVALNITKLSAIKTKYLREL 104  
Qy 119 -IVESTSESQILMMESRSKVDVAVKFEKSSSSSSSSSTKPEASDTAKPNKPTPEG 177  
Db 105 NVLEKSKDELPA---SEIKALDAAFEKFKDT-----LKPQ 138  
Qy 178 EKVAEAKKKVEAEKKAQDQKEEDRRNYPTITYKTLEIAESDVVKAELELVKVKAN 237  
Db 139 EKVAEAKKKVEAEKKAEDQKEEDRRNYPTNTYKTLEIAEAFDVVKAELELVKEAK 198  
Qy 238 EPREQIKIQAEEVSKQAEATRLKKITDRBAEAEAKRRADAK-----EQKP 289  
Db 199 ESRNEGITKQAEKVESKAEATLENIKTRDKAEAEAKKADAKLKEANVATSDQGP 258  
Qy 289 KGRAKGVPGELATPDKKENDAKSSDSVGBETLPSLSLPEKKVBAEKKVBAEKKA 348  
Db 259 KGRAKGVPGELATPDKKENDAKSSDSVGBETLPSLSLGGKKVBAEKKVBAEKKA 318  
Qy 349 DQKEEDRRNYPTNTYKTLEIAESDVVKAELELVKEAEKPRNEEKVKQAKAEVSK 408  
Db 319 DQKEEDRRNYPTNTYKTLDLSIASESDVKVKEAELELVKEAEKPRDEEIKIQAQAKVESK 378  
Qy 409 KAAATRLKIKITDRKKAEEAEKKAEDKVKEPAEQPAPAPKAEKPAPAPKAPENPA 468  
Db 379 KAAATRLKIKITDRKKAEEAEKKAEDKVKEPAEQPAPAPAPAPAPAPAPAPAP 436  
Qy 469 EQPAEKAPADQAEEDYARRSEYENRLTQQPPKTEKPAQSTP 513  
Db 437 EQPAEKAKTDDQAEEDYARRSEYENRLTQQPPKTEKPAQSTP 481

RESULT 3  
A97887  
surface protein pepA precursor [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97887  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: A97887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-619 <KUR>  
A:Cross-references: UNIPROT:Q54972; UNIPROT:Q8DR10; GB:AE007317; PIDN:AAK98925.1; PID:gl  
C:Genetics:  
A:Gene: pspA

Query Match 19.5%; Score 496; DB 2; Length 619;  
Best Local Similarity 33.7%; Pred. No. 3.2e-12;  
Matches 168; Conservative 62; Mismatches 148; Indels 120; Gaps 21;

Qy 26 VASLVMSVGHATENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKI 85  
Db 14 VALLGAGFVASQPTVVRAESPV-----ASQSKAEKDYDAAKDAK--NAKKAVEDAQKAL 67  
Qy 86 VGSYAKSTKKRHTITVALVNE---LNNIKNYLNKIVESTSLSQILMMESRSKVD 142  
Db 68 ---DDAKAAQKYDEDOQKTEEKAALKAASEEMDKAVAAVQQAYL-----AYQQA 115  
Qy 143 VSKFEKSSSSSSSSSTKPEAS-----DTAKPNKPTPEGKVAEAKKKVEAEKKAQDQ 198  
Db 116 TDKAAKDAADKIDEAKKREBEAKTKFNTVRMVVPEP-EQLAETKKSEAKQK----- 170  
Qy 199 EEDRRNYPTITYKTLEIAESDVVKAELELVKVKANEPREDEQIKIQAEEVSKQAE 258  
Db 171 -----PELT-KKLEBAKAEAEKKAQKQVDAEVAPOAKIAELENQVHRLQE 222  
Qy 259 ATRIKKITDRBAEAEAKRRADAKGEQKPKGRAKGVPGELATPDKKENDAKSSDSV 318  
Db 223 ---LKEI--DESESEDYAKGEFRAPLQSK-----LDAAKAKLS-- 255  
Qy 319 EETLPSLSLPEKKVBAEKKVBAEKKAQDQKEEDRRNYPTNTYKTLEIAESDVVK 378  
Db 256 -----KLEELSKIDE-----LDALIAKLELDQK 279  
Qy 379 KAELELVKEAEKPRN-EKVKQA-KAEVSKKAEATRLKIKITDRKKAEEAEKKA-AE 435  
Db 280 AAE-----ENNVEDYFKGLEKTIKAAKAE--LEKTEADLAKVNEPEKPA 327  
Qy 436 EDKVKEPAPQAPAPAPKAEKPAPAPKPNPAEQPAEKAPADQAEEDYARRSEYEN 495  
Db 328 ETPAPEAPAEQPKPAPAPAP---QAPAPKPEKPAEQPKPEKTDQQAEDYARRSEYEN 384  
Qy 496 LTQOQPKTEKPA-QPST 512  
Db 385 LTQOQPPKAEKPAKPAPT 402

RESULT 4  
A41971  
surface protein pepA precursor - Streptococcus pneumoniae  
N:Alternate names: pneumococcal surface protein A  
C:Species: Streptococcus pneumoniae  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41971; A60282; A33134  
R:Yother, J.; Briles, D.E.  
J. Bacteriol. 174, 601-609, 1992  
A:Title: Structural properties and evolutionary relationships of PepA, a surface protein  
A:Reference number: A41971; MUID:92105030; PMID:1729249





A;Residues: 1-1110 <AC>  
A;Cross-references: UNIPROT.Q91255; EMBL.U19361; NID:G632548; PIDN.AAA80106.1; PID:G6325E  
C;Superfamily: neurofilament triplet H protein

Query Match                  15.5%    Score 393.5 : DB 2 ; Length 1110 ;  
Best Local Similarity       26.3%; Pred.No. 3.8e-08 ;  
Matches 147 ; Conservative 106 ; Mismatches 229 ; Indels 77 ; Gaps 21 ;

Qy	3	ASKSERKVHYSIKFSYGVASVVVVASLVMSGVVHATENE--GATQVTPTSSNRANESQAEG	60
Dd	435	AAKVEKKV---VSKPKPIKESEPISAQLDITLEDLAQEENVWEAKAAPVPVAEKDEEEEEE	492
Qy	61	GEOPKKLDLS-----ERDKARKEVEEVVKIIGVESYAKSTKKRHTTIVALVNELNIKN	114
Dd	493	ESEEKEEAABAEEDDRCKGEAEAEBEAEBEVEKEEAABAEBEA---BAETETAFAE	549
Qy	115	YLNKIVESTSESOLQI---LMMESRSKVDNAVSKFEKDSSSSS--SSSDSTTKPEASDTAK	169
Dd	550	AAEEEEBAEGEEBAGABGAEBEGAEBEVBESAIEKAEEAEEAKAVEBESEAEEEEEEAE	609
Qy	170	PNNPTEPGCVBAKKVVEAAFKKADKKOEEDRRNYPTII---VKTLLEIAESDVVEVK	226
Dd	610	BEEVBAETKEVEZAAEAEBEGBEAAAEBEAAAAEAEVEVS KKAKTQEA VEVEEEAEAAE	669
Qy	227	AELEL-----VKKANEPREDOKIKOABAVESKOAEATRUKKI-KTDREAAEE	275
Dd	670	AEAEAEAEBAEBEDVBAESKE-EEBDSKEADAEDAEEADEAEFEEKGEEVTVKSDAEBAEA	728
Qy	276	KRRRAAKEQGKPGRAGRGPVELATPDKKENDAKSSDSVGEEITLPSPSLKPEKKVAE	335
Dd	729	AESEA-AKSEEEAAEAADBAEEBAEAAAEEAAETEATAEEAFAKEA--SDDEKPBEEVKE	785
Qy	336	AEKKVBEAKKAEDQKEEDEDRRYNTTYKTLELIAESDVVEVKKAELELVKEAKEPRNE	395
Dd	786	SEAPVAPEAKAPEKPAAPPKKAPAK----VESPTSPEDEPK---AEVVEKGKAEAPK	838
Qy	396	EKVQAKAFVESKAAETRUEKITRKKAEBEBAKRKAAEEDKVGEKPAEQ-OPAPAPK	454
Dd	839	PKAKPAAAKKEAKPV-----KESEPESPTEEPPKPPAA-----AKPAKAPAKPKAPK	888
Qy	455	A---EXPAPA----PKPENAEQPKAEPKPADQAQAEEDY-----ARSSEE	492
Dd	889	AEAEKEPEPAKPAQAAPAAAEDEDEDDEDEEVEEVKPEDAKPVKSKPAPAKEEE	948
Qy	493	YNRLTOQQPKTB-KPAQP 510	
Dd	949	DEPKAPAQFPKPKRPARP 967	

RESULT 7  
FCSOAG  
IGA FC receptor precursor - Streptococcus agalactiae  
N;Alternate names: beta antigen  
C;Species: Streptococcus agalactiae  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: S15330; S20240; S17038  
R;Jernstroem, P.G.; Chhatwal, G.S.; Timmis, K.N.  
Mol. Microbiol., 5, 843-849, 1991  
A;Title: The Iga-binding beta antigen of the c protein complex of Group B streptococci:  
A;Reference number: S15330; MUID:91312121; PMID:1857207  
A;Accession: S15330  
A;Molecule type: DNA  
A;Residues: 1-1164 <JE1>  
A;Cross-references: UNIPROT.P27951; EMBL.X59771  
A;Accession: S20240  
A;Molecule type: protein  
A;Residues: 38-48 <JE2>  
R;Jernstroem, P.G.  
submitted to the EMBL Data Library, August 1991  
A;Reference number: S17038  
A;Accession: S17038  
A;Molecule type: DNA  
A;Residues: 1-914,'E','916-1164 <JE3>

A;Cross-references: EMBL:X59771; NID:G46522; PIDN:CAA42442.1; PID:G46523  
C;Superfamily: IGA FC receptor  
C;Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;38-1164/Product: IGA FC receptor #status experimental <MAT>  
F;199-438/Domain: IGA binding #status predicted <IGA1>  
F;439-826/Domain: IGA binding #status predicted <IGA2>  
F;827-945/Region: proline-rich repeats  
F;946-1131/Domain: cell wall-spanning #status predicted <CWS>  
F;1132-1159/Domain: transmembrane #status predicted <TMM>

Query Match 14.4%; Score 365.5; DB 1; Length 1164;  
Best Local Similarity 28.2%; Pred. No. 4.5e-07;  
Matches 143; Conservative 81; Mismatches 180; Indels 103; Gaps 19;

Qy 1 MFASKSRKVHYSIRKFSVGVASVIVASVLMGSSVHATNEGATQVPTSSNRA---NES 56  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 1 MFKNGYERKMYSIIRKFSVGVASVAVASVLMGSSVAHSELVKDQSVKTTVEAAKPYPSMA 60  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 57 QAEQEQPKKLDSRDKAR-----KEVEYVKIVGESYAKSTKKRHTITVALVNLNN 110  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 61 QTDQGNSSSSSELETTKMEIPTTDIKKAVPEVKTAGETSATDTGKEK---QLQWKNN 117  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 111 IKNEYLNKIVSTSSQILMMESRKVDKAVSKPEKSSSSSSSSSSSSSTTKPEASDTAKP 170  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 118 LKNDVDNTILSHEQKN-----EFKTKIDE-----TNSDALLELE----- 152  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 171 NKPTPEGKVAEAKKKVEAEKKAKDKEEDRRNYPTITYKTLELEIAESDVEV--KKA 228  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 153 NQFNETNRLHIIKHVEEVEKKKAKQKK-----TLKQSDTKVDLSNIDKELNHQSK 204  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 229 LELVKVKANEPDEQIKQAEAEVSKQAEATRLKKIKTDREAEAEKERRADA-----K 283  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 205 VE-----KWAQKGITNEKDSMLKKIEDIRKQA-QQADKKEDAEVKKVRE 248  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 284 EOGKPKGKAKGVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKKVAEAKKVEEA 343  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 249 ELGKLFSTKAGLDQEIQEHVKKTSSEENTOKVDHYANSLQNLAKSLEELDKAT--- 305  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 344 KKAEDQKEEDRRNYPTITYKTLELE--IAESDVEVKA---EELVKKEAEKPRNEE-- 396  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 306 ---TNEQATQVKNOFLENAQKLKEIPLIKETNVKLYKAMSSELSQVEKELK--HNS 360  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 397 -----KVKQAKAEVSKAEATRLKIKTDREAEAEKAKR---AAEDKVKKEKPAEQ 447  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 361 LEDLVAKSKEIVREYEGKLNQSKNLPKL-----KQLEEEAHSKLKQVVEDPRKKFKTSE 416  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 448 QPAPAPKAEKPAPAKPENPAEQPKAE 474  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 417 TPKKRVKRDLLA-----NENNOQKIE 437  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
RESULT 8  
QFHUH  
N;filament triplet H protein - human  
N;Alternate names: neurofilament protein, 112K  
C;Species: Homo sapiens (man)  
C;Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 09-Jul-2004  
C;Accession: S00979  
R;Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.  
EMBO J. 7, 1947-1955, 1988  
A;Title: The structure and organization of the human heavy neurofilament subunit (NF-H)  
A;Reference numbers: S00979; MUID:89328981; PMID:3138108  
A;Accession: S00979  
A;Molecule type: DNA  
A;Residues: 1-1020 <LEE>  
A;Cross-references: UNIPROT:P12036; EMBL:X15306; NID:G35028; PIDN:CAA33366.1; PID:G184143  
A;Note: it is uncertain whether Met-1 or Met-2 is the initiator  
C;Genetics:  
A;Gene: GDB:NEFH  
A;Cross-references: GDB:120225; OMIM:162230  
A;Map position: 22q12.1-22q13.1  
A;Introns: 295/1; 361/3; 403/2

A;Cross-references: EMBL:X59771; NID:G46522; PIDN:CAA42442.1; PID:G46523  
C;Superfamily: IGA FC receptor  
C;Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;38-1164/Product: IGA FC receptor #status experimental <MAT>  
F;199-438/Domain: IGA binding #status predicted <IGA1>  
F;439-826/Domain: IGA binding #status predicted <IGA2>  
F;827-945/Region: proline-rich repeats  
F;946-1131/Domain: cell wall-spanning #status predicted <CWS>  
F;1132-1159/Domain: transmembrane #status predicted <TMM>

Query Match 14.4%; Score 365.5; DB 1; Length 1164;  
Best Local Similarity 28.2%; Pred. No. 4.5e-07;  
Matches 143; Conservative 81; Mismatches 180; Indels 103; Gaps 19;

Qy 1 MFASKSRKVHYSIRKFSVGVASVIVASVLMGSSVHATNEGATQVPTSSNRA---NES 56  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 1 MFKNGYERKMYSIIRKFSVGVASVAVASVLMGSSVAHSELVKDQSVKTTVEAAKPYPSMA 60  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 57 QAEQEQPKKLDSRDKAR-----KEVEYVKIVGESYAKSTKKRHTITVALVNLNN 110  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 61 QTDQGNSSSSSELETTKMEIPTTDIKKAVPEVKTAGETSATDTGKEK---QLQWKNN 117  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 111 IKNEYLNKIVSTSSQILMMESRKVDKAVSKPEKSSSSSSSSSSSSSTTKPEASDTAKP 170  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 118 LKNDVDNTILSHEQKN-----EFKTKIDE-----TNSDALLELE----- 152  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 171 NKPTPEGKVAEAKKKVEAEKKAKDKEEDRRNYPTITYKTLELEIAESDVEV--KKA 228  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 153 NQFNETNRLHIIKHVEEVEKKKAKQKK-----TLKQSDTKVDLSNIDKELNHQSK 204  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 229 LELVKVKANEPDEQIKQAEAEVSKQAEATRLKKIKTDREAEAEKERRADA-----K 283  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 205 VE-----KWAQKGITNEKDSMLKKIEDIRKQA-QQADKKEDAQVKVRE 248  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 284 EOGKPKGKAKGVPGELATPDKKENDAKSSDSVSGEETLPSPSLKPEKKVAEAKKVEEA 343  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 249 ELGKLFSTKAGLDQEIQEHVKKETSSBENTOKVDEHYANSLQNLAKSLEELDKAT-- 305  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 344 KKAEDQKEEDRRNYPTITYKTLELE--IAESDVEVKA---EELVKKEAEKPRNEE-- 396  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 306 ---TNEQATQVKNQFLENAQKLKEIPLIKETNVKLYKAMSSELEQVEKELK--HNS 360  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 397 -----KVKQAKAEVSKAEATRLKIKTDREAEAEKAKR---AAEDKVKKEKPAEQ 447  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 361 LEDLVAKSKEIVREYEGKLNQSKNLPQL-----KQLEEAHSHKLVQVVEPRKKFKTSE 416  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 448 QPAPAPKAEKPAPAKPENPAEQPKAE 474  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Qy 417 TPKKRVKRDLLA-----NENNOQKIE 437  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
RESULT 8  
QFHUH  
N;filament triplet H protein - human  
N;Alternate names: neurofilament protein, 112K  
C;Species: Homo sapiens (man)  
C;Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 09-Jul-2004  
C;Accession: S00979  
R;Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.  
EMBO J. 7, 1947-1955, 1988  
A;Title: The structure and organization of the human heavy neurofilament subunit (NF-H)  
A;Reference numbers: S00979; MUID:89328981; PMID:3138108  
A;Accession: S00979  
A;Molecule type: DNA  
A;Residues: 1-1020 <LEE>  
A;Cross-references: UNIPROT:P12036; EMBL:X15306; NID:G35028; PIDN:CAA33366.1; PID:G184143  
A;Note: it is uncertain whether Met-1 or Met-2 is the initiator  
C;Genetics:  
A;Gene: GDB:NEFH  
A;Cross-references: GDB:120225; OMIM:162230  
A;Map position: 22q12.1-22q13.1  
A;Introns: 295/1; 361/3; 403/2

C:Superfamily: neurofilament triplet H protein  
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein  
F:1-100/Domain: amino-terminal <NTD>  
F:101-410/Domain: rod #status predicted <ROD>  
F:411-1020/Domain: carboxyl-terminal <CTD>  
F:502-826/Region: 14-residue repeats  
F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,646  
(covalent) #status predicted  
F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 14.3%; Score 365; DB 1; Length 1020;  
Best Local Similarity 25.6%; Pred. No. 4.2e-07;  
Matches 150; Conservative 77; Mismatches 247; Indels 112; Gaps 20;  
  
QY 5 KSERKVH--SIRKPSVGASVVVASLVGVSVVHATENEGATQVPTSSNRANESQAEQGE 62  
DB KSEIKIVKVESEKTVIVEEQTEETQVTEEVTEEEKEKEEKGEGGEEBEEAEGE 498  
  
QY 63 QPKLDSERDKA--RKEVEEYK-KIVGESYAKSTKKRHTITVALVNLNLIKNEYLNKI 119  
DB EETKSPPAEEAASPEKEAKSPVKEEAKSPAEEAKSPEKEAKSPAEEAKSPVKEAKSP 553  
  
QY 120 VESTSESQILQMLMESRSKVDKAVSKFEKDSVSSSSSSSTKPEASDTAKPNKPT-EPGE 178  
DB AKEEAKSPPEAKSPKEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 613  
  
QY 179 KVAEAKKVVEAKKAKQKEDRRNYPTITYKTLEIAESDVEVKAELELVKVKANE 238  
DB SPAEAKSPVKEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKS 662  
  
QY 239 PRDEQIKIQAEAFVSK-----QAEATRLIKITD-REAEAEAKRRADAKQGGPKGRA 292  
DB PEKAKSPVKEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 722  
  
QY 293 KRGVPGELATPDKKENDAKSDSSVGEETLPSPLKPEKVAEAEKKVVEAKKAEADQKE 352  
DB KSPVKEEAKTPKAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 775  
  
QY 353 EDRRNYPTITYKTLEIAESD-----VEYKAELELV-KVEEAKSPNEE-----KV 398  
DB SPADKPEKAKSPVKEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAK 835  
  
QY 399 QKAEVSKKAEATRLKIKITDRKAEAEAKRKAEDKVEK----- 442  
DB KEPPKKAEEKAPATPKTEKKDSKK--EAPKKEAPKPVKEEKEPAPVKEPKSKVEAK 893  
  
QY 443 --PAEQPAPAPAKAEKP-----APAKPENPAEQPAPAEKP 476  
DB KEEAEKKKVPTPEKAEAPKVEKEDAKPEKTEVAKPEDDAKAEKSPKPAEKKEA-AP 952  
  
QY 477 ADQQAEEDYARRSEBYNRLTQQQPPKTEKPA-----QPSTP 513  
DB EKQDTKEEAKKPEK-----PKTEAKKEDDKTLSEKPSKP 989

RESULT 9  
T09127  
Iga Fc receptor precursor - Streptococcus agalactiae (strain SB35)  
C:Species: Streptococcus agalactiae  
C:Date: 11-Jun-1999 #sequence revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09127  
R:Heden, L.O.; Frithz, E.; Lindahl, G.  
Proc. Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998  
A:Title: A family of chimeric erythrocyte binding proteins of malaria parasites.  
A:Reference number: Z16577; MUID:98115903; PMID:9448314  
A:Accession: T09127  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1701 <KAP>  
A:Cross-references: EMBL:AF031886; NID:g2947227; PID:g2947228  
A:Experimental source: subspecies yoellii; strain YM  
C:Genetics:  
A:Gene: maebl

A:Introns: 62/1; 1648/1; 1674/2; 1697/1  
C:Keywords: alternative splicing; cell binding; erythrocyte invasion  
Query Match 14.2%; Score 361.5; DB 2; Length 1701;  
Best Local Similarity 29.6%; Pred. No. 9e-07;  
Matches 142; Conservative 75; Mismatches 168; Indels 95; Gaps 19;  
  
QY 57 QAEQGEQPKLDSERDKAR--KEVEEYKIVGESYAKSTKKRHTITVALVNLNLI--- 111  
DB EAQAAWAKAEAEERKKAEEAVKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1166  
  
QY 112 -KNEVLNKIVESTESQILQMLMESRSKVDKAVSKFEKDSVSSSSSSSTKPEASDTAKP 170  
DB KAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1205  
  
QY 171 NKPTPEGEKVAAEKVKVEAEKKAQ--DQKEDRRNYPTITYKTLE-----LEIAESDVEVK 225  
DB KAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1264  
  
QY 226 KAELELVKVKANEPDRDEQIKIQAE--EAEVESKQAEATRLKIKITDRKAEAEERKRA 283  
DB KAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1306  
  
QY 284 EQGPKGPAKGVPGELATPDKKENDAKSDSSVGEETLPSPLKPEKVAEAEKKVVEA 343  
DB LQRPKPKKGRSGKESRRRKKGRLKQKQK-----KKKKRLKQKQKQK 1353  
  
QY 344 KKAEDQKEDRRNYPTITYKTLEIAESDVEVKAELELVKVEAEKPRNEEYK---Q 400  
DB KRLKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1412  
  
QY 401 KAEVESKKAATRLKIKITDRKAEAEAKR-----KAAEDK-----VKEPAPQOPAP 451  
DB KAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1463  
  
QY 452 APKAEKAPAPKPNPAEQPAPKAEKPAQQAEE-----DYARRSEBYNRLTQQQPPKTEK 506  
DB EKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAEKAEAEERKTEAE 1521  
  
RESULT 10  
A60234  
Iga Fc receptor precursor - Streptococcus agalactiae (strain SB35)  
N:Alternate names: Iga-binding protein; protein Bac  
N:Contains: beta antigen  
C:Species: Streptococcus agalactiae  
C:Date: 08-Dec-1992 #sequence revision 08-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A60234; S14595; A60230  
R:Heden, L.O.; Frithz, E.; Lindahl, G.  
Eur. J. Immunol. 21, 1481-1490, 1991  
A:Title: Molecular characterization of an Iga receptor from group B streptococci: sequen  
ents with Iga-binding capacity.  
A:Reference number: A60234; MUID:91257158; PMID:2044657  
A:Accession: A60234  
A:Molecule type: DNA  
A:Residues: 1-1134 <HED>  
A:Cross-references: UNIPROT:Q99051; EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521  
A:Note: the source is designated as group B streptococcus strain SB35  
R:Heden, L.; Frithz, E.; Lindahl, G.  
Submitted to the EMBL Data Library, March 1991  
A:Description: Molecular characterization of an Iga receptor from group B streptococci: s  
fragments.  
A:Reference number: S14595  
A:Accession: S14595  
A:Molecule type: DNA  
A:Residues: 1-1134 <HE2>  
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521  
A:Note: the source is designated as Streptococcus agalactiae  
R:Lindahl, G.; Akersstrom, B.; Vaerman, J.P.; Stenberg, L.  
Eur. J. Immunol. 20, 2241-2247, 1990  
A:Title: Characterization of an Iga receptor from group B streptococci: specificity for s  
A:Reference number: A60230; MUID:91055597; PMID:2242758  
A:Accession: A60230





```
QY 429 AK-----RKAABEDVK--EKPAQOPAPAPAKPAKPAKPENPAEQPKA 473
DB 1823 LKSPTKKEKSPSTTKTKGDSKESPEKPEKPK-SPTPK-KSPPGSPK-KKSKSPEA 1879
QY 474 EKPADQQAED 484
DB 1880 EKPPAPKLTRD 1890

RESULT 15
S51364
sperm tail-specific protein mst101(2) - fruit fly (Drosophila hydei)
C:Species: Drosophila hydei
C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: S51364; S34154
R:Neesen, J.; Padmanabhan, S.; Buenemann, H.
Eur. J. Biochem. 225, 1089-1095, 1994
A:Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represe
lpha-helical rods within the extremely elongated spermatozoa of Drosophila hydei.
A:Reference number: S51364; MUID:95045538; PMID:7957199
A:Accession: S51364
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1390 <NEE>
A:Cross-references: UNIPROT:Q08696; EMBL:X73481
R:Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A:Reference number: S34153
A:Accession: S34154
A:Molecule type: DNA
A:Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>
A:Cross-references: EMBL:X73481; NID:G313201; PID:G313202
C:Genetics:
A:Gene: mst101(2)
A:Cross-references: FlyBase:FBgn0011816

Query Match 12.9%; Score 329.5; DB 2; Length 1390;
Best Local Similarity 26.5%; Pred. No. 1.2e-05;
Matches 137; Conservative 81; Mismatches 224; Indels 75; Gaps 14;

QY 39 ENEGATQVPTSSNRANE-SQAEQGEQPKLDSER---DKARKEVEEYVKIVGSEYAKST 94
DB 362 EKKACKELAKKKKKADEKKECEEAANKKKAERKKCKEKAERKEKAAEKKKCEAAKKE 421
QY 95 KKRHTITVALVNLNNIKNEYLNKIVBSTSBSQILMMESRSKVDKAVSKFEDSSSS 154
DB 422 KE-----AAERKKCEELAKNKKAAEKK---KCKEAAKKEKAAERKKCEELAKK 468
QY 155 SSDSSTKPEASDTAKPNKPTPEGKVAPAKKKVBEAE-----KKAKDQKE----- 199
DB 469 IKKAAEKKKCBETAKGKGEVAERKKCEELAKKIKKASIKKKCKKLAKKEREKTAEEKKCEK 528
QY 200 EDRNYPITITVKTLELETAESDVEVKAELELVKVAKEPRDEQIKQA-----EAEVES 254
DB 529 AAKRKGAAEKKCKEAKKKKKEAAEKKCKEKSAAKKEAAERKKCEAAKKEKAAEKK 588
QY 255 KQABATRLKIKITDREAAEEAEKRRADAKQCKPKGRAK-----RGVPGELATPDCKE 307
DB 589 KCEEAARKEVEAERKKCEELAKKIKKAAEKKCKEAAKKEKAAEAREKCGELAKKIKKA 648
QY 308 NDAKSSSSSVGEETLPSPSLKPE-----KKVAEAKKVBEA--KKKAEDQKEE 353
DB 649 AEKKKKCKLAKKEREKTAEEKKCEAAKKEAAEKKCKEKAEEKKCAEAAKKEAAEKKCEAAKK 708
QY 354 DRRNYPNTYKTLLELETAESDVEVKAELELVKVAKEPRNEEKVQAKAEVESKKAAT 413
DB 709 EKEAAERKKCEELAKKTK-----KAAEKKCKKLAKKKKAGKGNKLGKNGKGA--- 759
QY 414 RLEKIKTRKKAEEAKRAAEDVKVEKPAEQPAPAPAKPAKPAKPENPAEQPKA 473
DB 760 -----LKEKKCKRELAKKGAEEKKCKEAAKKEKAAEKKCKEKTAKKKEE--AEKKKC 812
QY 474 EKPADQQAEDYARRSEE-EYNRLTQQQPPKTEKPAQ 509
```

```
DB 813 EKTAKGKEAAEKKKKCKEKAAGKGRKEAEKKCKEKTAK 849

Search completed: September 28, 2005, 18:43:32
Job time : 49 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2005, 18:29:20 ; Search time 171 Seconds  
(without alignment)

1160.281 Million cell updates/sec

Title: US-09-298-523D-13

Perfect score: 2546

Sequence: 1 MFASKSRKHVHSRKPSVG.....NRLTQQQPPKTEKPAQFSTP 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_l6Dec04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2546	100.0	693	5	AAU76890 Pneumococ
2	2546	100.0	693	6	ABU02720 S. pneumo
3	2546	100.0	693	8	ADM92271 S. pneumon
4	2546	100.0	694	3	AAy81653 Streptoco
5	2366	92.9	655	2	AAy32099 Choline b
6	2366	92.9	655	2	AAy49226 CbpA of s
7	2366	92.9	655	7	ABW01575 Streptoco
8	2366	92.9	655	8	ADRI0795 Choline b
9	2268	89.1	460	2	AAy49250 N-termina
10	2268	89.1	460	8	ADRI0831 Choline b
11	2265	89.0	460	2	AAy32189 N-termina
12	2251	88.4	459	2	AAy32190 N-termina
13	2251	88.4	459	2	AAy49251 N-termina
14	2251	88.4	459	8	ADRI0832 Choline b
15	2210	86.8	446	2	AAy49143 Amino aci
16	2193	86.1	446	2	AAy49140 Amino aci
17	2117	83.2	428	2	AAy32110 Choline b
18	2117	83.2	428	2	AAy49238 Choline b
19	2117	83.2	428	7	ABW01587 Streptoco
20	2117	83.2	428	8	ADRI0817 Choline b
21	1995	78.4	406	2	AAy32098 Choline b
22	1995	78.4	406	2	AAy49225 Polypepti
23	1995	78.4	406	7	ABW01574 Streptoco
24	1995	78.4	406	8	ADRI0794 Choline b
25	1786.5	70.6	487	5	AAU76889 Pneumococ

26	1756	69.0	564	2	AAW62654	AAW62654 C3 bindin
27	1756	69.0	701	6	ABU46304	ABU46304 Protein e
28	1752	68.8	701	5	AAU76888	AAU76888 Pneumococ
29	1744	68.5	581	2	AAy43394	AAy43394 S. pneumo
30	1653	64.9	437	8	ADRI0833	ADRI0833 Choline b
31	1635.5	64.2	488	5	AAU76887	AAU76887 Pneumococ
32	1581	62.1	663	2	AAy32104	AAy32104 Choline b
33	1581	62.1	663	2	AAy43395	AAy43395 S. pneumo
34	1581	62.1	663	2	AAy49232	AAy49232 CbpA of s
35	1581	62.1	663	5	AAE29201	AAE29201 Streptoco
36	1581	62.1	663	5	ABG60639	ABG60639 Partial s
37	1581	62.1	663	6	ABP55319	ABP55319 Streptoco
38	1581	62.1	663	7	ABW01581	ABW01581 Streptoco
39	1581	62.1	663	7	ADL99571	ADL99571 GST-piGR
40	1581	62.1	663	8	ADRI0801	ADRI0801 Choline b
41	1580	62.1	419	2	AAy32182	AAy32182 N-termina
42	1580	62.1	419	2	AAy49243	AAy49243 N-termina
43	1580	62.1	419	8	ADRI0824	ADRI0824 Choline b
44	1524	59.9	406	2	AAy49152	AAy49152 Amino aci
45	1517.5	59.6	631	2	AAW46444	AAW46444 CbpA, a c

## ALIGNMENTS

RESULT 1  
AAU76890  
ID AAU76890 standard; protein; 693 AA.  
XX  
AC AAU76890;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pneumococcal surface protein C19 TIGR (PspC19 TIGR).  
XX  
KW Hic; factor H; fH; antibacterial; PspC19 TIGR;  
KW pneumococcal surface protein C19 TIGR.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200208426-A2.  
XX  
PD 31-JAN-2002.  
XX  
PF 19-JUL-2001; 2001WO-EP008409.  
XX  
PR 20-JUL-2000; 2000SE-00002738.  
XX  
PA (HANS-) HANSA MEDICAL AB.  
XX  
PI Bjorck L, Sjöholm A, Janulczyk R, Pozzi G, Iannelli F;  
XX  
DR WPI; 2002-257337/30.  
XX  
PT Polypeptide having ability to bind factor H, and proteins and peptides  
PT derived from polypeptides useful as vaccine for infections caused by  
PT bacteria and for identifying agents inhibiting binding of factor H to  
protein.  
XX  
PS Disclosure; Page 59-61; 63pp; English.  
XX  
CC The invention relates to the Streptococcus pneumoniae Hic polypeptide  
CC having the ability to bind factor H (fH). The Hic protein and the  
CC polynucleotide encoding it can be used to make vaccine compositions  
CC capable of generating an immune response to Streptococcus pneumoniae or  
CC binding to an anti-protein Hic antibody. The protein and its homologues  
CC are useful for identifying an agent which inhibits binding of factor H to  
CC Streptococcus pneumoniae comprising incubating any one of the  
CC polypeptides with factor H and a test agent, monitoring binding of fH to  
CC the polypeptide and determining thereby whether the test agent inhibits  
CC binding of factor to the polypeptide. This sequence represents the  
CC pneumococcal surface protein C19 TIGR (PspC19 TIGR), a factor H binding  
CC protein

```
XX SQ Sequence 693 AA;
Query Match 100.0%; Score 2546; DB 5; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.8e-152;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFASKSERKVHYSIRKFSVGASVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MFASKSERKVHYSIRKFSVGASVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60

Qy 61 GEOPKLDSDRKARKEVEYVKIVGESYAKSTKGRHTTITVALVNLNINKNEYLNKIV 120
Db 61 GEOPKLDSDRKARKEVEYVKIVGESYAKSTKGRHTTITVALVNLNINKNEYLNKIV 120

Qy 121 ESTSESQILIMMESRSKVDPAVSKFPEKDSOSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180
Db 121 ESTSESQILIMMESRSKVDPAVSKFPEKDSOSSSSSSDSTKPEASDTAKPNKPTPEGEKV 180

Qy 181 AEAKKVVEAEKKAQKQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240
Db 181 AEAKKVVEAEKKAQKQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240

Qy 241 DEQIKIQAEAEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRKRGVPGEL 300
Db 241 DEQIKIQAEAEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRKRGVPGEL 300

Qy 301 ATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKKAEDQKEEDRRNYP 360
Db 301 ATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKKAEDQKEEDRRNYP 360

Qy 361 NTYKTLLEIAESDVEVKAELELVKEAEKPRNEEKVKQAKAEVESKKAETRLKIKT 420
Db 361 NTYKTLLEIAESDVEVKAELELVKEAEKPRNEEKVKQAKAEVESKKAETRLKIKT 420

Qy 421 DRKXAEAEAKKAAEDDKVKEKPAEQPPAPAPKAEKPAPAPKPPENPAEQPKAEKPADQQ 480
Db 421 DRKXAEAEAKKAAEDDKVKEKPAEQPPAPAPKAEKPAPAPKPPENPAEQPKAEKPADQQ 480

Qy 481 AEDYARRSEEVNRLTQQQPPKTEKPAQSTP 513
Db 481 AEDYARRSEEVNRLTQQQPPKTEKPAQSTP 513
```

```
RESULT 2
ABU02720
ID ABU02720 standard; protein; 693 AA.
XX AC ABU02720;
XX XX 23-OCT-2003 (revised)
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #2299.
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
XX KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
XX KW gene therapy; vaccine.
XX OS Streptococcus pneumoniae; type 4 strain.
XX PN WO200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-IB002163.
XX XX 27-MAR-2001; 2001GB-00007658.
XX PR (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Masignani V, Tettelin H, Fraser C;
```

```
XX DR WPI: 2003-040579/03.
XX N-PSDB; ABX08011.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX ear infection.
XX PS Claim 1; SEQ ID NO 4598; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX SQ Sequence 693 AA;
```

```
Query Match 100.0%; Score 2546; DB 6; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.8e-152;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFASKSERKVHYSIRKFSVGASVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MFASKSERKVHYSIRKFSVGASVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60

Qy 61 GEOPKLDSDRKARKEVEYVKIVGESYAKSTKGRHTTITVALVNLNINKNEYLNKIV 120
Db 61 GEOPKLDSDRKARKEVEYVKIVGESYAKSTKGRHTTITVALVNLNINKNEYLNKIV 120

Qy 121 ESTSESQILIMMESRSKVDPAVSKFPEKDSOSSSSDSTKPEASDTAKPNKPTPEGEKV 180
Db 121 ESTSESQILIMMESRSKVDPAVSKFPEKDSOSSSSDSTKPEASDTAKPNKPTPEGEKV 180

Qy 181 AEAKKVVEAEKKAQKQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240
Db 181 AEAKKVVEAEKKAQKQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240

Qy 241 DEQIKIQAEAEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRKRGVPGEL 300
Db 241 DEQIKIQAEAEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRKRGVPGEL 300

Qy 301 ATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKKAEDQKEEDRRNYP 360
Db 301 ATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKKAEDQKEEDRRNYP 360

Qy 361 NTYKTLLEIAESDVEVKAELELVKEAEKPRNEEKVKQAKAEVESKKAETRLKIKT 420
```





CC agent is useful for treatment or prophylaxis of *S. pneumoniae* infection  
CC and meningitis. AAA05591 to AAA05614 represent primers used in the  
CC exemplification of the present invention  
XX  
SQ Sequence 694 AA;

Query Match 100.0%; Score 2546; DB 3; Length 694;  
Best Local Similarity 100.0%; Pred. No. 1.8e-152;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFASKSRKVHYSIRKSVGVASVWVSLVMSVGHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSRKVHYSIRKSVGVASVWVSLVMSVGHATENEGATQVPTSSNRANESQAEQ 60  
Qy 61 GEQPKLDSERDKARKEVEEVKKIVGESYAKSTKKRHTIIVLVNLLNINKYLNKIV 120  
Db 61 GEQPKLDSERDKARKEVEEVKKIVGESYAKSTKKRHTIIVLVNLLNINKYLNKIV 120  
Qy 121 ESTSESQILMMESRSKVDPAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGEKV 180  
Db 121 ESTSESQILMMESRSKVDPAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGEKV 180  
Qy 181 AEAKKVEEAKKADQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240  
Db 181 AEAKKVEEAKKADQKEEDRRNYPITYTKTLEIAESDVEVKAELELVKVKANEPR 240  
Qy 241 DEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRAGVPGEL 300  
Db 241 DEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKRRADAKSQGPKGRAGVPGEL 300  
Qy 301 ATPDKKENDAKSSDSSVGEETLPSPLSKPEKKVAAEAKKVEAEKKAEDQKEEDRRNYP 360  
Db 301 ATPDKKENDAKSSDSSVGEETLPSPLSKPEKKVAAEAKKVEAEKKAEDQKEEDRRNYP 360  
Qy 361 NTYKTLEIAESDVEVKAELELVKBEAKPRNEEKVKQAEVESKKAATRLKIKT 420  
Db 361 NTYKTLEIAESDVEVKAELELVKBEAKPRNEEKVKQAEVESKKAATRLKIKT 420  
Qy 421 DRKAAEAEAKKAAEDKVKKEKPAEQOPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 480  
Db 421 DRKAAEAEAKKAAEDKVKKEKPAEQOPAPAPKAEKPAPAPKAPENPAEQPKAEKPADQ 480  
Qy 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513  
Db 481 AEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513

RESULT 5  
AAV32099  
ID AAV32099 standard; protein; 655 AA.  
XX  
AC AAV32099;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Choline binding protein A (Cbpa).  
XX  
KW Choline binding protein; Cbpa; adhesin; immunogen; vaccine; diagnosis;  
KW therapy; pneumococcus; sepsis; meningitis; otitis media; pneumonia.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO9951188-A2.  
XX  
PD 14-OCT-1999.  
XX  
PF 07-APR-1999; 99WO-US007669.  
XX  
PR 07-APR-1998; 98US-00056019.  
PR 07-APR-1998; 98US-0080878P.  
XX  
XX (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
FA (MEDI-) MEDIMMUNE INC.

XX  
PI Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;  
XX WPI; 1999-633690/54.  
DR N-PSDB; AAZ34452.  
DR  
XX  
PT New N-terminal choline binding protein A truncate polypeptides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections.  
XX  
PS Disclosure; Page 88-93; 160pp; English.

XX This sequence represents the choline binding protein A (Cbpa) of  
CC Streptococcus pneumoniae serotype type 4. The invention provides novel N-  
CC terminal Cbpa truncated polypeptides (see AAY32098-110 and AAY32179-90)  
CC and polynucleotides encoding them, host-vector systems, and antibodies  
CC that specifically bind to the polypeptides. The invention also relates to  
CC vaccines including the polypeptides, which provide protection or elicit  
CC protective antibodies to bacterial infection, specifically pneumococcus,  
CC and to antibodies and antagonists against such polypeptides for use in  
CC diagnosis and passive immunotherapy. The polypeptides and/or  
CC polynucleotides are also useful as competitive inhibitors of bacterial  
CC adhesion of pneumococcus  
XX  
SQ Sequence 655 AA;

Query Match 92.9%; Score 2366; DB 2; Length 655;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 39 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEVKKIVGESYAKSTKKR 98  
Db 1 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEVKKIVGESYAKSTKKR 60  
Qy 99 TITVALVNLNINKYLNKIVESTSESQILMMESRSKVDPAVSKFEKDSSSSSSSDS 158  
Db 61 TITVALVNLNINKYLNKIVESTSESQILMMESRSKVDPAVSKFEKDSSSSSSSDS 120  
Qy 159 STPEASDTAKPNKPTPEGEKVAAEAKKVEAEKKAEDQKEEDRRNYPITYTKTLEIEIA 218  
Db 121 STPEASDTAKPNKPTPEGEKVAAEAKKVEAEKKAEDQKEEDRRNYPITYTKTLEIEIA 180  
Qy 219 ESDVEVKAELELVKVKANEPRDEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKR 278  
Db 181 ESDVEVKAELELVKVKANEPRDEQIKIQAEEVESKQAEATRLKKIKTDREAEAEAKR 240  
Qy 279 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPLSKPEKKVAAEK 338  
Db 241 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPLSKPEKKVAAEK 300  
Qy 339 KVEEAKKKAEDQKEEDRRNYPITYTKTLEIEIAESDVEVKAELELVKBEAKPRNEEKV 398  
Db 301 KVEEAKKKAEDQKEEDRRNYPITYTKTLEIEIAESDVEVKAELELVKBEAKPRNEEKV 360  
Qy 399 KQAEAVESKKAATRLKIKTDREAEAEAKRKAABEDKVKKEKPAEQOPAPAPKAEKP 458  
Db 361 KQAEAVESKKAATRLKIKTDREAEAEAKRKAABEDKVKKEKPAEQOPAPAPKAEKP 420  
Qy 459 APAPKPNPAEQPKAEKPADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 513  
Db 421 APAPKPNPAEQPKAEKPADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPSTP 475

RESULT 6  
AAV49226  
ID AAV49226 standard; protein; 655 AA.  
XX  
AC AAV49226;  
XX  
DT 07-FEB-2000 (first entry)  
XX  
DE Cbpa of serotype 4 amino acid sequence.  
XX

KW Choline binding protein A; CbpA; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
OS Streptococcus pneumoniae.  
PN WO9951187-A2.  
XX 14-OCT-1999:  
XX 07-APR-1999; 99WO-US007669.  
XX 07-APR-1998; 98US-00056019.  
PR 07-APR-1998; 98US-0080878P.  
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA Tuomanen EI, Masure HR;  
XX WPI; 1999-620161/53.  
DR N-PSDB; AAZ31402.  
XX Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections.  
XX Disclosure; Page 9; 85pp; English.  
XX The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (CbpA) truncate  
CC (CbpAT). The polypeptides can be selected from sequences shown in  
CC AAY49225, AAY49227, AAY49230, AAY49231, AAY49233 and AAY49238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use  
XX  
SQ Sequence 655 AA;  
Query Match 92.9%; Score 2366; DB 2; Length 655;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 39 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKKIIGESYAKSTKKRH 98  
DB 1 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKKIIGESYAKSTKKRH 60  
QY 99 TITVALNELNLIKNEYLNKIVESTSESQQLILMMESRKVDEAVSKFEKSSSSSDS 158  
DB 61 TITVALNELNLIKNEYLNKIVESTSESQQLILMMESRKVDEAVSKFEKSSSSSDS 120  
QY 159 STKPEASTAKPNKTEPEGEKVAEAKKVEEAKKQKQEDRRNYPTITVKTLELEIA 218  
DB 121 STKPEASTAKPNKTEPEGEKVAEAKKVEEAKKQKQEDRRNYPTITVKTLELEIA 180  
QY 219 ESDVVEKKAELVLVKANEPDEQIKQAEAEVESKQAEATRLKKIKTDREAEAEAKR 278  
DB 181 ESDVVEKKAELVLVKANEPDEQIKQAEAEVESKQAEATRLKKIKTDREAEAEAKR 240  
QY 279 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKVAEAK 338  
DB 241 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKVAEAK 300  
QY 339 KVEEAKKAEDQKEEDRRNPNTYKTLLELEIAESDVEVKKAELVLKVEEAKEPNEESKV 398  
DB 301 KVEEAKKAEDQKEEDRRNPNTYKTLLELEIAESDVEVKKAELVLKVEEAKEPNEESKV 360

QY 399 KOAKAEVESKKAETRLKIKTDREKKAEEBAKRAAEEDKVKKEKPAEQQPAPAPKAEKP 458  
DB 361 KOAKAEVESKKAETRLKIKTDREKKAEEBAKRAAEEDKVKKEKPAEQQPAPAPKAEKP 420  
QY 459 APAPKPNPAPQKAEKPADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPSSTP 513  
DB 421 APAPKPNPAPQKAEKPADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPSSTP 475  
RESULT 7  
ABW01575  
ID ABW01575 standard; protein; 655 AA.  
XX AC ABW01575;  
XX 12-FEB-2004 (first entry)  
XX Streptococcus pneumoniae cbpA serotype 4 protein.  
XX Therapy; choline-binding protein A; cbpA; immune response; immunisation;  
KW therapeutic; diagnosis; vaccine; antibacterial.  
XX Streptococcus pneumoniae.  
XX FH Key Location/Qualifiers  
FT Misc-difference 344 /note= "Encoded by GCTT"  
XX US2003096950-A1.  
XX 22-MAY-2003.  
XX 07-APR-1998; 98US-00056019.  
XX 07-APR-1998; 98US-00056019.  
XX (TUOM/) TUOMANEN E I.  
PA (WIZE/) WIZEMANN T M.  
PA (MASU/) MASURE H R.  
PA (JOHN/) JOHNSON L S.  
PA (KOEN/) KOENIG S.  
PI Tuomanen EI, Wizemann TM, Masure HR, Johnson LS, Koenig S;  
XX WPI; 2003-843081/78.  
DR N-PSDB; AAD62881.  
XX New N-terminal fragment of choline-binding protein A, useful for  
PT treatment or prevention of infection by Streptococcus pneumoniae, also  
PT new nucleic acid.  
XX Disclosure; Page 3; Opp; English.  
XX The present invention provides novel isolated polypeptide comprising an  
CC amino acid sequence of an N-terminal choline-binding protein A (cbpA)  
CC truncate. The invention is useful for induction of a protective or  
CC therapeutic immune response against pneumococci (or other cross-reactive  
CC bacteria). Antibodies raised against pneumococci are useful for passive  
CC immunisation and diagnosis. The invention is also useful as vaccines. The  
CC present sequence is Streptococcus pneumoniae choline-binding protein A  
CC (cbpA) serotype 4 protein  
XX Sequence 655 AA;  
Query Match 92.9%; Score 2366; DB 7; Length 655;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 39 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKKIIGESYAKSTKKRH 98  
DB 1 ENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKKIIGESYAKSTKKRH 60



XX  
PI Tuomanen EI, Masure HR;  
XX WPI; 1999-620161/53.  
XX Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections.  
XX  
XX Disclosure; Fig 2A-B; 85pp; English.  
XX  
XX The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC AAY49225, AAY49227, AAY49230, AAY49231, AAY49233 and AAY49238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use  
XX  
XX Sequence 460 AA;  
XX  
XX Query Match 89.1%; Score 2268; DB 2; Length 460;  
XX Best Local Similarity 99.8%; Pred. No. 4.1e-135;  
XX Matches 458; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 25 VVASLWVGSVVHATENEGATOVPTSSNRANESQAEGQPKLDSERDKARKEVEEYVK 84  
DB 2 IVASLWVGSVVHATENEGATOVPTSSNRANESQAEGQPKLDSERDKARKEVEEYVK 61  
QY 85 IVGESYAKSTKKRHTITVALVNLNINIKVEYLNKIVESTSSQLQILMWESRSKYDEAVS 144  
DB 62 IVGESYAKSTKKRHTITVALVNLNINIKVEYLNKIVESTSSQLQILMWESRSKYDEAVS 121  
QY 145 KFEKDSRSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKAKDKQEEEDRN 204  
DB 122 KFEKDSRSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKAKDKQEEEDRN 181  
QY 205 YPTITYKTLELEIAESDVEVKAELELVKVKANEPDEQIKQAEAEVESKQAEATRLKK 264  
DB 182 YPTITYKTLELEIAESDVEVKAELELVKVKANEPDEQIKQAEAEVESKQAEATRLKK 241  
QY 265 IKTDREAEAEAKRRADAKAQKPKGRKRGVPGELATPDKKENDAKSSDSSVSGEETLPS 324  
DB 242 IKTDREAEAEAKRRADAKAQKPKGRKRGVPGELATPDKKENDAKSSDSSVSGEETLPS 301  
QY 325 PSLKPEKKVAEAEKVKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 384  
DB 302 PSLKPEKKVAEAEKVKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 361  
QY 385 VKEAEKPRNEEKVKQAEAEVESKQAEATRLKIKTDREAEAEAKRRADAKAQKPKGRK 444  
DB 362 VKEAEKPRNEEKVKQAEAEVESKQAEATRLKIKTDREAEAEAKRRADAKAQKPKGRK 421  
QY 445 EOQAPAPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 483  
DB 422 EOQAPAPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 460  
RESULT 10  
ADRI0831  
ID ADRI0831 standard; protein; 460 AA.  
XX  
AC ADRI0831;  
XX

DT 23-SEP-2004 (first entry)  
XX Choline binding protein A (Cbpa) N-terminal region Ntype4.  
XX  
XX vaccine; pneumococcal infection; Streptococcus pneumoniae;  
KW choline binding protein A; Cbpa.  
XX Streptococcus pneumoniae.  
XX  
XX US2004120966-A1.  
XX  
XX 24-JUN-2004.  
XX  
XX 05-JAN-2004; 2004US-00751702.  
XX  
XX 07-APR-1998; 98US-00056019.  
XX (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX (MEDI-) MEDIMUNE INC.  
XX  
XX Tuomanen EI, Wizemann TM, Masure HR, Johnson LS, Koenig S;  
XX WPI; 2004-479616/45.  
XX  
XX New vaccine comprises a polypeptide, which does not bind to choline and  
PT exhibits a tertiary structure as found in a native, full-length Cbpa  
PT polypeptide, useful for treating or protecting against pneumococcal  
PT infection.  
XX  
XX Disclosure; SEQ ID NO 38; 70pp; English.  
XX  
XX The invention relates to a vaccine for treating or protecting against  
CC pneumococcal infection. The vaccine comprises a polypeptide in a  
CC pharmaceutical carrier where the polypeptide does not bind to choline,  
CC exhibits a tertiary structure as found in a native, full-length Cbpa  
CC polypeptide, and the polypeptide content of the vaccine is useful for  
CC treating or protecting against pneumococcal infection. The vaccine is useful for  
CC treating or protecting against pneumococcal infection, by inhibiting the  
CC adhesion of Pneumococcus, particularly Streptococcus pneumoniae. The  
CC present sequence represents the amino acid sequence of a choline binding  
XX protein A (Cbpa) N-terminal region.  
XX  
XX Sequence 460 AA;  
XX  
XX Query Match 89.1%; Score 2268; DB 8; Length 460;  
XX Best Local Similarity 99.8%; Pred. No. 4.1e-135;  
XX Matches 458; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 25 VVASLWVGSVVHATENEGATOVPTSSNRANESQAEGQPKLDSERDKARKEVEEYVK 84  
DB 2 IVASLWVGSVVHATENEGATOVPTSSNRANESQAEGQPKLDSERDKARKEVEEYVK 61  
QY 85 IVGESYAKSTKKRHTITVALVNLNINIKVEYLNKIVESTSSQLQILMWESRSKYDEAVS 144  
DB 62 IVGESYAKSTKKRHTITVALVNLNINIKVEYLNKIVESTSSQLQILMWESRSKYDEAVS 121  
QY 145 KFEKDSRSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKAKDKQEEEDRN 204  
DB 122 KFEKDSRSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVBEAEKAKDKQEEEDRN 181  
QY 205 YPTITYKTLELEIAESDVEVKAELELVKVKANEPDEQIKQAEAEVESKQAEATRLKK 264  
DB 182 YPTITYKTLELEIAESDVEVKAELELVKVKANEPDEQIKQAEAEVESKQAEATRLKK 241  
QY 265 IKTDREAEAEAKRRADAKAQKPKGRKRGVPGELATPDKKENDAKSSDSSVSGEETLPS 324  
DB 242 IKTDREAEAEAKRRADAKAQKPKGRKRGVPGELATPDKKENDAKSSDSSVSGEETLPS 301  
QY 325 PSLKPEKKVAEAEKVKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 384  
DB 302 PSLKPEKKVAEAEKVKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 361  
QY 385 VKEAEKPRNEEKVKQAEAEVESKQAEATRLKIKTDREAEAEAKRRADAKAQKPKGRK 444

```

Db      362 VKBEAKPRNEEKVKQAEVSKAEATRLLEKI KTDKKAEEBAKRAAEEDKVKSKPA 421
QY      445 EOQOPAPAPKAEPAPAPKPNPAEQPKAEKPADQQAEE 483
Db      422 EOQOPAPAPKAEPAPAPKPNPAEQPKAEKPADQQAEE 460

RESULT 11
AAY32189
ID AAY32189 standard; protein; 460 AA.
AC AAY32189;
XX
DT 01-FEB-2000 (first entry)
DE N-terminal choline binding protein A (Cbpa) truncate.
XX
KW Choline binding protein; Cbpa; truncate; adhesin; immunogen; vaccine;
KW diagnosis; therapy; pneumococcus; sepsis; meningitis; otitis media;
KW pneumonia.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
FN WO9951188-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US007669.
XX
PR 07-APR-1998; 98US-00056019.
PR 07-APR-1998; 98US-0080878P.
XX
PA (SUJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMUNE INC.
XX
PI Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;
XX WPI; 1999-633690/54.
XX
DR New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections.
XX
PS Claim 47; Fig 2A-B; 160pp; English.
XX
CC The present sequence represents an N-terminal choline binding protein A
CC (Cbpa) truncate, denoted Ntype 4 Cbpa trun, derived from Streptococcus
CC pneumoniae Cbpa. Claimed vaccines contain and N-terminal Cbpa truncate
CC such as the present sequence, or a polypeptide comprising a conserved
CC region of the Cbpa truncate. The vaccines provide protection or elicit
CC protective antibodies to bacterial infection, specifically pneumococcus.
CC Antibodies and antagonists against the N-terminal Cbpa truncates are
XX used in diagnosis and passive immunotherapy
XX
SQ Sequence 460 AA;

Query Match      89.0%; Score 2265; DB 2; Length 460;
Best Local Similarity 99.6%; Pred. No. 6.4e-135;
Matches 457; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      25 VVASLWVGSVVHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEYVKK 84
Db      2 IVASLWVGSVVHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEYVKK 61
QY      85 IVGSYAKSTKKRHTITVALVNLNINIKNEYLKIVESTSESQILMMSRSKVDKAVS 144
Db      62 IVGSYAKSTKKRHTITVALVNLNINIKNEYLKIVESTSESQILMMSRSKVDKAVS 121
QY      145 KFEKDSSSSSSDSTKPEASDTAKPNKPTPGKVAEAKKKAEKKAKDQKEEDRN 204
Db      122 KFEKDSSSSSSDSTKPEASDTAKPNKPTPGKVAEAKKKAEKKAKDQKEEDRN 181

```

```

QY      205 YPTITYKTLEIAESDVEVKAELELVKVKANEPRDEQKIKQAEAEVESKQAEATRLKK 264
Db      182 YPTITYKTLEIAESDVEVKAELELVKVKANEPRDEQKIKQAEAEVESKQAEATRLKK 241
QY      265 IKTDREBAEEAKRADAKQGGPKGRAKGVPGELATPDKKENDAKSSSSVGEETLPS 324
Db      242 IKTDREBAEEAKRADAKQGGPKGRAKGVPGELATPDKKENDAKSSSSVGEETLPS 301
QY      325 PSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPNTYKTLEIAESDVEVKAELEL 384
Db      302 PSLKPEKKVAEAKKVEEAKKAEADQKEEDRRNYPNTYKTLEIAESDVEVKAELEL 361
QY      385 VKBEAKPRNEEKVKQAEVSKAEATRLLEKI KTDKKAEEBAKRAAEEDKVKSKPA 444
Db      362 VKBEAKPRNEEKVKQAEVSKAEATRLLEKI KTDKKAEEBAKRAAEEDKVKSKPA 421
QY      445 EOQOPAPAPKAEPAPAPKPNPAEQPKAEKPADQQAEE 483
Db      422 EOQOPAPAPKAEPAPAPKPNPAEQPKAEKPADQQAEE 460

```

## RESULT 12

AAY32190  
ID AAY32190 standard; protein; 459 AA.

AC AAY32190;

XX  
DT 01-FEB-2000 (first entry)

XX N-terminal choline binding protein A (Cbpa) truncate.

XX Choline binding protein; Cbpa; truncate; adhesin; immunogen; vaccine;  
KW diagnosis; therapy; pneumococcus; sepsis; meningitis; otitis media;  
KW pneumonia.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO9951188-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-US007669.

XX 07-APR-1998; 98US-00056019.

PR 07-APR-1998; 98US-0080878P.

XX (SUJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PA (MEDI-) MEDIMUNE INC.

PI Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;

XX WPI; 1999-633690/54.

XX New N-terminal choline binding protein A truncate polypeptides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections.

XX Claim 47; Fig 2A-B; 160pp; English.

XX The present sequence represents an N-terminal choline binding protein A  
CC (Cbpa) truncate, denoted ATCC4 Cbpa trun, derived from Streptococcus  
CC pneumoniae Cbpa. Claimed vaccines contain and N-terminal Cbpa truncate  
CC such as the present sequence, or a polypeptide comprising a conserved  
CC region of the Cbpa truncate. The vaccines provide protection or elicit  
CC protective antibodies to bacterial infection, specifically pneumococcus.  
CC Antibodies and antagonists against the N-terminal Cbpa truncates are  
XX used in diagnosis and passive immunotherapy

SQ Sequence 459 AA;

Query Match

88.4%; Score 2251; DB 2; Length 459;

Best Local Similarity 98.9%; Pred. No. 4.9e-134;  
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 25 VVASLVMSGVVHATNEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 84  
:|||||  
Db 1 IVASLVMSGVVHATNEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 60  
:|||||

QY 85 IVGESYAKSTKKRHTITVALVNLNIIKNEYLKIVESTESQLOILMMESRSKYDEAVS 144  
:|||||  
Db 61 IVGESYAKSTKKRHTITVALVNLNIIKNEYLKIVESTESQLOILMMESRSKYDEAVS 120  
:|||||

QY 145 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKVEAEKAKDQKEEDRRN 204  
:|||||  
Db 121 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKVEAEKAKDQKEEDRRN 180  
:|||||

QY 205 YPTITYKTLELEIAESDVEVKKAELELVKVKANEPDRDQIKQAEAEVESKQAEATRLKK 264  
:|||||  
Db 181 YPTITYKTLELEIAESDVEVKKAELELVKVKANEPDRDQIKQAEAEVESKQAEATRLKK 240  
:|||||

QY 265 IKTDREAEAEAKRRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGETLPS 324  
:|||||  
Db 241 IKTDREAEAEAKRRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGETLPS 300  
:|||||

QY 325 PSLKPEKKVAEAEKVEAEKAKKAEQKEDRRNPTNTYKTLELEIAESDVEVKKAELEL 384  
:|||||  
Db 301 PSLKPEKKVAEAEKVEAEKAKKAEQKEDRRNPTNTYKTLELEIAESDVEVKKAELEL 360  
:|||||

QY 385 VKEEAKPRNEEKVKQAKAEVESKKAATRLLEKIKTDRKKAEEAEAKRKAABEDKVKKEKPA 444  
:|||||  
Db 361 VKEEAKPRNEEKVKQAKAEVESKKAATRLLEKIKTDRKKAEEAEAKRKAABEDKVKKEKPA 420  
:|||||

QY 445 EQQPAPAPKAPKAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 483  
:|||||  
Db 421 EQQPAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 459  
:|||||

RESULT 13  
AAY49251  
ID AAY49251 standard; protein; 459 AA.  
AC AAY49251;  
XX  
XX  
XX 07-FEB-2000 (first entry)  
XX  
XX N-terminal region of CbpA polypeptide ATCC4.  
XX  
XX Choline binding protein A; CbpA; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO9951187-A2.  
PN  
PD 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US007668.  
XX  
XX 07-APR-1998; 98US-00056019.  
PR 07-APR-1998; 98US-0080878P.  
XX  
XX (SJD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX  
XX Tuomanen EI, Masure HR;  
PI  
XX  
XX WPI; 1999-620161/53.  
DR  
XX  
XX Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections.  
PT  
XX  
XX Disclosure; Fig 2A-B; 85pp; English.  
PS  
XX  
XX The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (CbpA) truncate  
CC

(CbpA). The polypeptides can be selected from sequences shown in  
CC AAY49225, AAY49227, AAY49230, AAY49231, AAY49233 and AAY49238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use  
XX  
SQ Sequence 459 AA;

Query Match 88.4%; Score 2251; DB 2; Length 459;  
Best Local Similarity 98.9%; Pred. No. 4.9e-134;  
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 25 VVASLVMSGVVHATNEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 84  
:|||||  
Db 1 IVASLVMSGVVHATNEGATQVPTSSNRANESQAEGQPKKLDSEDKARKEVEEYVKK 60  
:|||||

QY 85 IVGESYAKSTKKRHTITVALVNLNIIKNEYLKIVESTESQLOILMMESRSKYDEAVS 144  
:|||||  
Db 61 IVGESYAKSTKKRHTITVALVNLNIIKNEYLKIVESTESQLOILMMESRSKYDEAVS 120  
:|||||

QY 145 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKVEAEKAKDQKEEDRRN 204  
:|||||  
Db 121 KFEKDSSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKVEAEKAKDQKEEDRRN 180  
:|||||

QY 205 YPTITYKTLELEIAESDVEVKKAELELVKVKANEPDRDQIKQAEAEVESKQAEATRLKK 264  
:|||||  
Db 181 YPTITYKTLELEIAESDVEVKKAELELVKVKANEPDRDQIKQAEAEVESKQAEATRLKK 240  
:|||||

QY 265 IKTDREAEAEAKRRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGETLPS 324  
:|||||  
Db 241 IKTDREAEAEAKRRADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGETLPS 300  
:|||||

QY 325 PSLKPEKKVAEAEKVEAEKAKKAEQKEDRRNPTNTYKTLELEIAESDVEVKKAELEL 384  
:|||||  
Db 301 PSLKPEKKVAEAEKVEAEKAKKAEQKEDRRNPTNTYKTLELEIAESDVEVKKAELEL 360  
:|||||

QY 385 VKEEAKPRNEEKVKQAKAEVESKKAATRLLEKIKTDRKKAEEAEAKRKAABEDKVKKEKPA 444  
:|||||  
Db 361 VKEEAKPRNEEKVKQAKAEVESKKAATRLLEKIKTDRKKAEEAEAKRKAABEDKVKKEKPA 420  
:|||||

QY 445 EQQPAPAPKAPKAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 483  
:|||||  
Db 421 EQQPAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 459  
:|||||

RESULT 14  
ADRI0832  
ID ADRI0832 standard; protein; 459 AA.  
XX  
XX  
AC ADRI0832;  
XX  
XX 23-SEP-2004 (first entry)  
DT  
XX  
XX Choline binding protein A (CbpA) N-terminal region ATCC4.  
DE  
XX  
XX vaccine; pneumococcal infection; Streptococcus pneumoniae;  
KW choline binding protein A; CbpA.  
KW  
XX  
XX Streptococcus pneumoniae.  
OS  
XX  
XX US2004120966-A1.  
PN  
XX  
XX 24-JUN-2004.  
PD

XX 05-JAN-2004; 2004US-00751702.  
XX 07-APR-1998; 98US-00056019.  
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (MEDI-) MEDIMUNE INC.  
XX Tuomanen EI, Wizemann TM, Masure HR, Johnson LS, Koenig S;  
XX WPI; 2004-479616/45.  
XX New vaccine comprises a polypeptide, which does not bind to choline and  
XX exhibits a tertiary structure as found in a native, full-length CbpA  
PT polypeptide, useful for treating or protecting against pneumococcal  
PT infection.  
XX Disclosure; SEQ ID NO 39; 70pp; English.  
XX The invention relates to a vaccine for treating or protecting against  
CC pneumococcal infection. The vaccine comprises a polypeptide in a  
CC pharmaceutical carrier where the polypeptide does not bind to choline,  
CC exhibits a tertiary structure as found in a native, full-length CbpA  
CC polypeptide, and the polypeptide content of the vaccine is for treating  
CC or protecting against pneumococcal infection. The vaccine is useful for  
CC treating or protecting against pneumococcal infection, by inhibiting the  
CC adhesion of Pneumococcus, particularly Streptococcus pneumoniae. The  
CC present sequence represents the amino acid sequence of a choline binding  
CC protein A (CbpA) N-terminal region.  
XX SQ Sequence 459 AA;  
Query Match 88.4%; Score 2251; DB 8; Length 459;  
Best Local Similarity 98.9%; Pred. No. 4.9e-134;  
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 25 VVASLVGMSVVVHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEVVKK 84  
Db :|||||  
1 IVASLVGMSVVVHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEVVKK 60  
Qy 85 IVGESYAKSTKKRHTITVALVNLNINKNEYLKIVESTSESQILMMESRSKVDKAVS 144  
Db :|||||  
61 IVGESYAKSTKKRHTITVALVNLNINKNEYLKIVESTSESQILMMESRSKVDKAVS 120  
Qy 145 KFEKDSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKAKDKQKEEDRN 204  
Db :|||||  
121 KFEKDSSSSSDSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKAKDKQKEEDRN 180  
Qy 205 YPTITYKTLELEIAESDVEVKKAELELVKVKANPRDEQKIQAEEVESKQAEATRLKK 264  
Db :|||||  
181 YPTITYKTLELEIAESDVEVKKAELELVKVKANPRDKQKIQAEEVESKQAEATRLKK 240  
Qy 265 IKTDREAEBAKRADAKGEGKPKGRKRGVPGELATPDKENDAKSSDSSVGEETLPS 324  
Db :|||||  
241 IKTDREAEBAKRADAKGEGKPKGRKRGVPGELATPDKENDAKSSDSSVGEETLPS 300  
Qy 325 PSLKPEKVAEAKKVEAEKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVKKAELEL 384  
Db :|||||  
301 PSLKPEKVAEAKKVEAEKKAEDQKEEDRRNPTNTYKTLELEIAESDVEVKKAELEL 360  
Qy 385 VKEAKEPRNEEKVQKAEVESKKAETRLLEKIKTDKKAEEBAKKAFAEDKVKEKPA 444  
Db :|||||  
361 VKEAKEPRNEEKVQKAEVESKKAETRLLEKIKTDKKAEEBAKKAFAEDKVKEKPA 420  
Qy 445 EQQPAPAPKAEKAPAPKPNPAEQKAEKPADQQAEE 483  
Db :|||||  
421 EQQPAPAPKAEKAPAPKPNPAEQKAEKPADQQAEE 459  
RESULT 15  
AA49143  
ID AA49143 standard; protein; 446 AA.  
XX

AC AAY49143;  
XX 17-JAN-2000 (first entry)  
XX Amino acid sequence of choline-binding protein fragment #7.  
DE Truncated surface binding protein; alpha helix; choline binding protein;  
XX vaccine; invasive bacterial infection; otitis media; sepsis; meningitis;  
KW lobar pneumonia infection; antibody; immature immune system;  
KW immunocompromised.  
XX Streptococcus pneumoniae.  
OS  
XX WO9951266-A2.  
PN  
XX 14-OCT-1999.  
PD  
XX 06-APR-1999; 99WO-US007680.  
PF  
XX 07-APR-1998; 98US-0080878P.  
PR  
XX 15-MAY-1998; 98US-0085743P.  
XX (MEDI-) MEDIMUNE INC.  
PA  
XX Wizemann TM, Koenig S, Johnson LS;  
PI  
XX WPI; 1999-601465/51.  
DR  
DR N-PSDB; AAZ31083.  
XX New pneumococcal proteins useful as vaccines and for diagnosis of  
PT pneumococcal infections.  
XX Claim 10; Page 72-73; 98pp; English.  
XX AAY49137-Y49152 are amino acid sequences that are fragments of choline  
CC binding proteins (CBP). The fragments of the protein are the alpha helix  
CC forming parts of the CBPs from Streptococcus pneumoniae. The polypeptides  
CC do not contain the actual choline binding fragment. The polypeptides and  
CC the nucleotide sequences that encode them (AAZ31077-231092) are used in  
CC the invention, which relates to polypeptide truncates of a pneumococcal  
CC surface binding protein containing the highly conserved immunogenic alpha  
CC helical portion and no choline binding portion. The polypeptides are used  
CC as immunogens in a bacterial vaccine. The vaccine can be used for  
CC preventing (immunising) or treating invasive bacterial (especially  
CC pneumococcal) infections, especially otitis media (caused by  
CC S.pneumoniae), sepsis, meningitis and lobar pneumonia infections.  
CC Antibodies raised against the polypeptide are useful for detection,  
CC prevention (passive immunity) and treatment of S. pneumoniae infections.  
CC The vaccines are especially useful in immunocompromised patients, those  
CC with an immature immune system, or patients with an on going pneumococcal  
CC infection. The vaccine avoids unnecessary expense and provides broad  
CC protection against a range of pneumococcal serotypes and it produces an  
CC improved and enhanced effect in preventing bacterial infections  
XX SQ Sequence 446 AA;  
Query Match 86.8%; Score 2210; DB 2; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.8e-131;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 38 TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEVVKIIGVESYAKSTKKR 97  
Db 1 TENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEEVVKIIGVESYAKSTKKR 60  
Qy 98 HTTITVALVNLNINKNEYLKIVESTSESQILMMESRSKVDKAVSKPKDSSSSSSSS 157  
Db :|||||  
61 HTTITVALVNLNINKNEYLKIVESTSESQILMMESRSKVDKAVSKPKDSSSSSSSS 120  
Qy 158 SSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKKAEDQKEEDRRNPTITYKTLELEI 217  
Db :|||||  
121 SSTKPEASDTAKPNKPTPEGKVAEAKKKVEAEKKAEDQKEEDRRNPTITYKTLELEI 180  
Qy 218 AESDVEVKKAELELVKVKANPRDEQKIQAEEVESKQAEATRLKKIKTDREAEBAEK 277



Db	181	AESDVEVKAELELVKVKANSFRDSQIKQAAEVESKQAAEATRLUKTKTDREEAEEAAK	240
Qy	278	RRADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGSETTLPSPSLKPEKKVAEAE	337
Db	241	RRADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGSETTLPSPSLKPEKKVAEAE	300
Qy	338	KKVEEAKKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKKAELVKEEAKPEPRNEEK	397
Db	301	KKVEEAKKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKKAELVKEEAKPEPRNEEK	360
Qy	398	VKQAAEAVESKKAETRLLEKTKDRKKAEEBAKRVKAAEEDKVKEKPAEQPQAPAPKPAEK	457
Db	361	VKQAAEAVESKKAETRLLEKTKDRKKAEEBAKRVKAAEEDKVKEKPAEQPQAPAPKPAEK	420
Qy	458	PAPAPKPNPAEQPKAEPADQQAEE	483
Db	421	PAPAPKPNPAEQPKAEPADQQAEE	446

Search completed: September 28, 2005, 18:39:36  
Job time : 174 secs

**THIS PAGE IS BLANK**